VENÓMICA. MECANISMOS MOLECULARES Y EVOLUTIVOS DE LA DIVERSIFICACIÓN ESTRUCTURAL DE LA FAMILIA DE LAS DISINTEGRINAS

TESIS DOCTORAL DE PAULA JUÁREZ GÓMEZ

DIRIGIDA POR EL DR. JUAN JOSÉ CALVETE Y LA DRA. LIBIA SANZ

Instituto de Biomedicina de Valencia

D. Juan José Calvete Chornet, Doctor en Biología y Profesor de Investigación en el Instituto de Biomedicina de Valencia, Consejo Superior de Investigaciones Científicas (CSIC) y Dª Libia Sanz Sanz, Doctora en Química

CERTIFICAN:

Que **Paula Juárez Gómez**, licenciada en Biología por la Universitat de València, ha realizado bajo su dirección el trabajo de Tesis Doctoral que lleva por titulo "Venómica. Mecanismos moleculares y Evolutivos de la diversificación estructural de la familia de las disintegrinas".

En Valencia, a 9 de enero de 2007

Dr. Juan José Calvete Chornet
Dpto. de Proteómica Estructural
Instituto de Biomedicina de Valencia

Dra. Libia Sanz Sanz
Dpto. de Proteómica Estructural
Instituto de Biomedicina de Valencia

A mi familia y amig@s

AGRADECIMIENTOS/ ACKNOWLEDGEMENTS

Parece que ha llegado el momento de dar gracias a toda la gente que ha hecho posible que esta Tesis haya salido adelante… empezaré por mi familia, sin ellos no estaría aquí delante de todos defendiendo mi trabajo. He tenido en casa el mejor ejemplo de tenacidad, espíritu de superación e ilusión, valores muy importantes a la hora de trabajar en un laboratorio. Siempre me he sentido respaldada en mis decisiones y me han animado cuando han venido "vacas flacas", incluso cuando no entendían el porqué de mis quejas!. Mis padres han sido, y serán, un pilar en mi vida que junto a mi hermano Alex, tías, tío, primos y abuelas forman mi hogar. Me siento muy orgullosa y afortunada de la familia que tengo. Aquí aprovecho para dar las gracias especialmente a mi iaia Alejandra. Ella es mi ejemplo de bondad y coraje. Te echo de menos. Espero llegar a alcanzar lo que ella deseaba para mí. Os quiero.

Si ahora tuviera que nombrar a todas las personas que han estado implicadas en esta andadura no acabaría nunca pero empezare por Manolo Portolés e Inma Azorín del Centro de Investigación de La Fe que fueron mi primer contacto con la poyata y me dieron el empujón que necesitaba para tomar la decisión crucial, pero muy acertada, de dar el salto al CSIC. Cuando doy las GRACIAS al IBV lo hago en mayúsculas porque tanto a nivel científico como personal este centro está por encima de la media, mejor dicho, se sale!. Gracias a las chicas de administración, las de la limpieza (Luisa y Susi que siempre me suben la moral!), Amparo, Paco, Nacho, Manolo, los de seguridad (que sería de mi sin Yolinda!), al Sr. Cubells y sus eventos y a nuestra Sra. Gerente. En cuanto al personal científico… que puedo decir! A lo largo de estos años creo que he usado algo de alguno de los 11 laboratorios del centro y a todos os doy las gracias por vuestra ayuda. En especial quiero agradecer al lab de la Dra. Casado por las PCRs, cuantificaciones, enzimas y consejos y al del Dr. Sanz por el electroporador y placas de última hora.

A nivel personal quiero dar las gracias a Juanjo y Libia, no sólo por los conocimientos científicos que he aprendido de ellos sino por introducirme el gusanillo de querer saber más y más. Me han tratado casi como una hija (ja, ja), incluso ha habido épocas que los he visto más que a mis verdaderos padres!. Si todo mi trabajo ha salido adelante a sido gracias también a Alicia. Ella ha sido mi confidente y mis manos en el labo cuando he estado fuera. Celso empezó este camino conmigo, hemos compartido risas y gruñidos. Vas a ser un Gran Jefe!. Miguel llegó justo a tiempo para equilibrar la guerra de sexos en el labo, aunque luego se ha decantado hacia los XY por la vuelta de Pepe y el último en llegar, Pedro, que estoy segura que va a disfrutar/aprender mucho. Toda la gente que ha pasado por mi labo ha dejado un poquito de ellos mismos en esta Tesis y a todos vosotros os doy las gracias.

Sabía que llegado el momento de daros las gracias a vosotros mis compañeros, y sin embargo amigos, me quedaría sin palabras. Son tantas emociones compartidas: alegría, tristeza, frustración, optimismo... Laia, nunca olvidaré las carreras por la 5ª Av., el margarita en el Village y las confesiones por el messenger. Santi, gracias por escucharme, pero también por aguantar mis silencios. Cintia, por mucho que te empeñes no me engañas, eres todo ternura y ahora te ha llegado el momento de demostrarlo!!. Hemos compartido muchas risas y alguna que otra lagrimita y aunque me ponga sentimental sabéis de sobra que os quiero y que siempre me tendréis para lo que queráis. Gracias a Sandra y Pili por hacerme participe de vuestra faceta de mamás, a Amalia (recuerda lo mucho que vales!), a Belén (viva Robbie!), mis chic@s del 3º que valen su peso en oro, en especial MC-Ada-Leda-Leo-Jose LL.-Fer-Jose G.-Mariano, vuestras sonrisas y algún que otro piropo me han animado más de un día gris. Gracias Marta por los achuchones y por tener tu puerta siempre abierta para mí. A los que ya no están por aquí: Mª Paz, Rafa y Kiko (espero que cumplas y cuando seas jefe me guardes un hueco en tu lab! Gracias por intentar comprenderme, tarea ardua y complicada!!).

Fuera de estas cuatro paredes esta mi "otro mundo". Mis amigos han sido mi válvula de escape a tanta ciencia, aunque mis BioCompis no han tenido escapatoria y en algún momento he tenido que tirar de su ayuda para mis experimentos. Gracias Txa, Silvi-David, Dani, Vicente-Sandra, Gsus-Alba, Pepe y Mer. ¡Menos mal que nos quedaban 4 días en el convento! Ja, ja. Gracias Juanma-Marisol, Ana, Luisito, Paula y Moni. Las que a la fuerza han acabado entendiendo que es eso del veneno de las serpientes han sido mis amigas Belén, Sara, Ampa, Amparo, Bárbara, Ilona (quina paciència mare!), Carmen, Irene, Lorelai, Luisa y Sara. Elisabet y Eva, habéis sido mi mejor apoyo y compañía en Liverpool'06. A toda la gente que he tenido la suerte de conocer en mis estancias (Philly'03 y Liverpool). Ellos hicieron de vivir en el extranjero una aventura personal además de profesional. Todos ellos han tenido que aguantar mis "rollos" más de una vez, así que ahora es el momento de decir GRACIAS de todo corazón.

Of course I would like to thank people from Venom Unit in The School of Tropical Medicine in Liverpool. Thanks Rob and Dave for giving me the chance to work in your lab and introducing me to the exciting world of Snake Venoms! I have enjoyed every minute I've spent with you. Even after our first meal together you wanted me in your lab! Thanks Si for all your brilliant knowledge about molecular biology and bioinformatics, without your help I couldn't have done it. I can't miss Jen, Gav, Daz, Paul, Helen and Fran, all of them were very helpful and cheered me up when things weren't going well. I can say very proudly that more than colleagues I've got good friends over there. Mary and Amanda's smiles have been a good reason to return to the Pool. I'd like to extend my gratitude to Kelly, Clare,

AGRADECIMIENTOS/ ACKNOWLEDGEMENTS

Gian, Claudia, Dee, Lee, Isabella, people from Dr. Bates' lab, especially Davina, the porters and all the DTMH'06 gang. To all of you, many many thanks!

¡GRACIAS A TOD@S!

Paula

Índice

1. INTRODUCCIÓN

1.1 LAS SERPIENTES

La inmensa mayoría de las cerca de 3000 especies de serpientes actuales no son venenosas, y -como se discutirá más adelante- los varios cientos de especies de ofidios potencialmente peligrosos para el hombre albergan en sus glándulas del veneno un arsenal químico, cuya composición y acción biológica han sido refinadas a lo largo de millones de años de evolución y que representa tanto un arma mortífera para la presa como una farmacopea natural cuyo enorme potencial biotecnológico y clínico está siendo activamente explorado en laboratorios de todo el orbe (Menez 2002; Menez et al. 2006).

Se estima que hace unos 310 millones de años los reptiles descendieron de los anfibios. La innovación evolutiva que permitió a los reptiles separarse de los anfibios y colonizar ecosistemas terrestres fue el desarrollo de huevos con cáscara que podían ser depositados en tierra, liberando a los reptiles de la necesidad de retornar al agua para reproducirse (http://www.geocities.com/CapeCanaveral/Hangar/2437/evolve.html) (Greene 1997). El registro fósil incluye ejemplares de serpientes que datan del periodo Cretácico, hace unos 135 millones de años, e indican el origen terrestre de las serpientes, cuya versión moderna son los boídos (boas, pitones y anacondas). Estas serpientes convivieron tímidamente con los dinosaurios, los reptiles que dominaron la Tierra hasta su desaparición hace unos 65 millones de años durante la transición entre los periodos Cretácico y Terciario (línea K-T), debido a las consecuencias climáticas del impacto de un asteroide en el Golfo de México (Alvarez et al. 1980). Entonces fue cuando los boídos pasaron a dominar el planeta. Actualmente se sitúa a las serpientes junto a los lagartos, con quienes comparten un ancestro común (Kochva 1987), en el orden Squamata de la clase Reptilia, subfilo Vertebrata, filo Chordata, del reino animal (Fig.1)

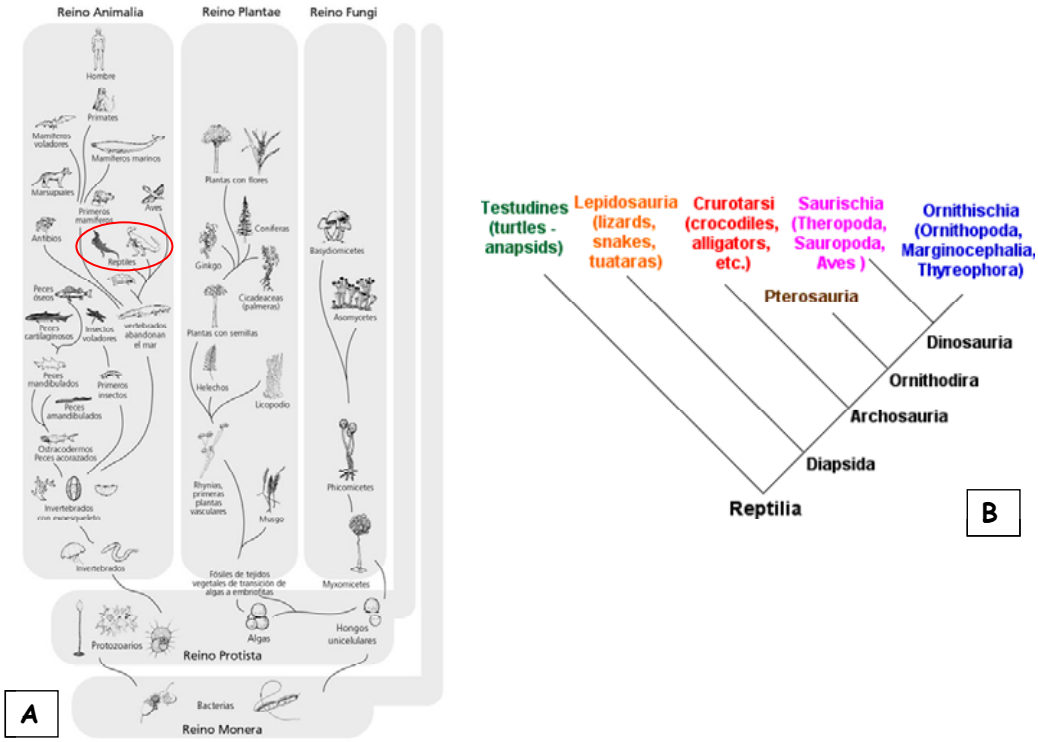


Figura 1. A) Esquema de los 5 Reinos de la naturaleza (Whittaker y Margulis 1978). Basado en la organización celular, complejidad estructural y modo de nutrición B) Cladograma de la clase Reptilia donde se muestran las diferentes subclases y los órdenes que la forman.

Hace unos 35 millones de años apareció un grupo de serpientes más pequeñas y rápidas, los colubroideos, que compitió con los boídos por comida y nichos ecológicos. Los colubroideos constituyeron un pequeño grupo de serpientes hasta hace unos 20 millones de años (Mioceno), cuando las placas tectónicas se alejaron del ecuador alcanzando sus posiciones actuales disminuyendo drásticamente las temperatura locales. Los boídos, incapaces de adaptarse a estos cambios climáticos, se vieron relegados de gran parte de sus nichos ecológicos, los cuales fueron colonizados rápidamente por los colubroideos que pronto dominaron el universo de las serpientes. Su radiación coincide con la aparición y diversificación de los roedores, mamíferos que constituyen una parte muy importante de la dieta en muchos de ellos

(http://galeon.hispavista.com/tartaret/aficiones1089995.html). Las aproximadamente 3000 especies actuales de serpientes se agrupan en unos 400 géneros y 18 familias (http://www.embl-heidelberg.de/~uetz/families/snakes.html), presentes en hábitats tanto terrestres como acuáticos, desde el mar hasta los desiertos, en todos los continentes, exceptuando la Antártida. El éxito evolutivo de los colubroideos se refleja en el hecho de que este grupo de serpientes incluye a más de dos tercios de todas las especies de ofidios actuales.

La mayor parte de los colubroideos son completamente inofensivos, ya que a pesar de poseer la glándula de Duvernoy activa en la producción de toxinas (Ching et al. 2006; Fry et al. 2003; Huang y Mackessy 2004) son aglifas (carecen de aparato inoculador) y su saliva no es lo bastante tóxica como para constituir un peligro para la presa. De hecho, aunque la aparición de la glándula del veneno data de hace unos 200 millones de años, durante la evolución de los reptiles escamosos (Fry et al. 2006), no es hasta hace unos 10-15 millones de años que diferentes grupos de serpientes desarrollaron independientemente aparatos de inyección del veneno (Jackson 2003). Las serpientes denominadas opistoglifas desarrollaron en uno de los dientes posteriores de cada mitad de la mandíbula superior, de mayor tamaño que los demás, un canal que facilita el paso del veneno. Este tipo de dentición es frecuente en los colúbridos. Los proteroglifos (elápidos) poseen uno o más colmillos en la parte anterior de los maxilares con un surco que puede estar cerrado en algún tramo. La abertura del surco permite a algunas cobras, como la cobra esputante *Naja sputatrix*, escupir su veneno relativamente lejos. Los solenoglifos (vipéridos y atractaspídidos) son los ofidios que poseen el sistema de inyección de veneno más elaborado. El colmillo es un diente muy largo y el canal de inyección está cerrado en toda su extensión, permitiendo que la inoculación del veneno sea profunda.

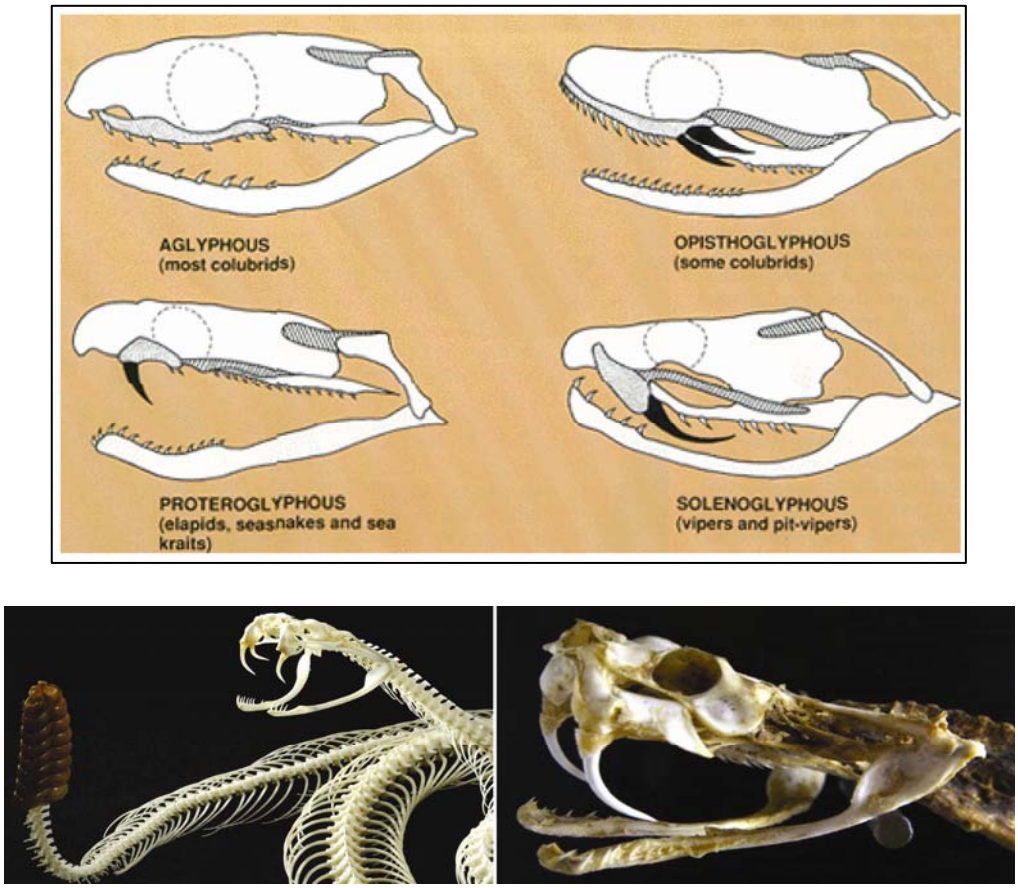


Figura 2. Panel superior, esquema de tipos de dentición de serpientes resaltando los colmillos que sirven para inyectar el veneno. Abajo, esqueleto de *Crotalus atrox* y cráneo de una serpiente de la familia Viperidae donde se observan los colmillos maxilares inyectores del veneno.

La evolución del aparato venenoso constituyó la adaptación clave que posibilitó a las serpientes venenosas la transición de un modo mecánico (constricción) a un modo químico (veneno) de matar e ingerir presas mucho más grandes que ellas mismas. Otras adaptaciones, como la de la cobra, sirven también como mecanismo de defensa.

1.2. COMPOSICIÓN DE LOS VENENOS DE SERPIENTES

En el mundo existen unas 3000 especies de serpientes de las que unas 640 son venenosas. Según la OMS anualmente ocurren 5,4 millones de accidentes por mordedura de serpientes, de los cuales 2,7 millones producen envenenamiento y dan lugar a más de 125.000 muertes al año (Theakston et al. 2003). Las especies de serpientes venenosas se clasifican esencialmente en dos grandes familias, Viperidae (víboras y serpientes de cascabel) y Elapidae (cobras, mambas, serpientes de coral, etc). Aunque los venenos son mezclas complejas de moléculas farmacológicamente activas cuyos efectos biológicos también son complejos debido a que los distintos componentes pueden actuar individual- o sinergicamente, atendiendo a su acción biológica primaria pueden clasificarse en neuro-/miotóxicos o hemorrágicos/citotóxicos. En el primer grupo se encuentran los venenos de la familia Elapidae, ricos en neurotoxinas (bloquean canales iónicos dependientes de voltaje presentes en el cerebro y en las uniones neuromusculares, produciendo una parálisis de la presa) y fosfolipasas A_2 de tipo I, cuyos efectos abarcan miotoxicidad, cardiotoxicidad, actividad pro- y anti-coagulante, así como neurotoxicidad pre- y post-sináptica. La mayoría de las serpientes australianas pertenecen a esta familia y son reconocidas como las más tóxicas del planeta (Fry 1999; Harvey 2001).

Los venenos de los vipéridos (subfamilias Viperinae y Crotalinae) poseen un arsenal de proteínas capaces de degradar la matríz extracelular e interfir con la cascada de coagulación, el sistema hemostático y la reparación de tejidos (Fox y Serrano 2005a; Kini 2006; Markland 1998). Las manifestaciones clínicas del envenenamiento por vipéridos y crotálidos pueden ser locales o sistémicas (Gutierrez et al. 2005; White 2005). Los efectos

locales se presentan minutos después de la inyección del veneno e incluyen con frecuencia dolor, edema, equimosis y hemorragia local. Tales signos son seguidos en muchos casos por necrosis del área que rodea al sitio de la mordedura. Entre los efectos de tipo sistémico se incluyen alteraciones en la coagulación sanguínea y episodios hemorrágicos distantes al sitio de inyección del veneno (Marsh 1994). Además, muchos venenos contienen toxinas que afectan al sistema cardiovascular, habiéndose caracterizado factores de crecimiento endotelial (svVEGF) que alteran la permeabilidad vascular (Suto et al. 2005) y péptidos que inhiben a la enzima convertidora de la angiotensina I en angiotensina II, potenciando las acciones biológicas de la bradiquinina. La consecuencia es una inmediata y severa bajada de la presión sanguínea (Hayashi y Camargo 2005; Joseph et al. 2004).

1.3. PROTEÍNAS DE VENENOS DE SERPIENTES DE LA FAMILIA VIPERIDAE

A pesar de la aparente complejidad de los venenos de víboras y serpientes de cascabel evidenciada mediante técnicas de separación de proteínas, como cromatografía de fase reversa y electroforesis bidimensional (Serrano et al. 2005), análisis detallados de los proteomas han revelado que las proteínas de estos venenos pertenecen tan solo a unas 10-12 familias (Bazaa et al. 2005; Juárez et al. 2004; Juárez et al. 2006a; Sanz et al. 2006b). Análisis de los transcriptomas de un número creciente de serpientes de la familia Viperidae (*Bitis gabonica*, *Bothrops insulares, Bothrops jararacussu*, *Agkistrodon acutus*, *Echis ocellatus* y *Lachesis muta*) corroboran esta noción (Francischetti et al. 2004; Junqueira-de-Azevedo Ide y Ho 2002; Junqueira-de-Azevedo et al. 2006; Kashima et al. 2004; Qinghua et al. 2006; Wagstaff y Harrison 2006). La composición proteica de los venenos refleja el hecho de que éstos se originaron en etapas tempranas de la evolución de los

reptiles escamosos por reclutamiento y transformación mediante evolución acelerada de un número reducido de proteínas endógenas (Fry et al. 2006).

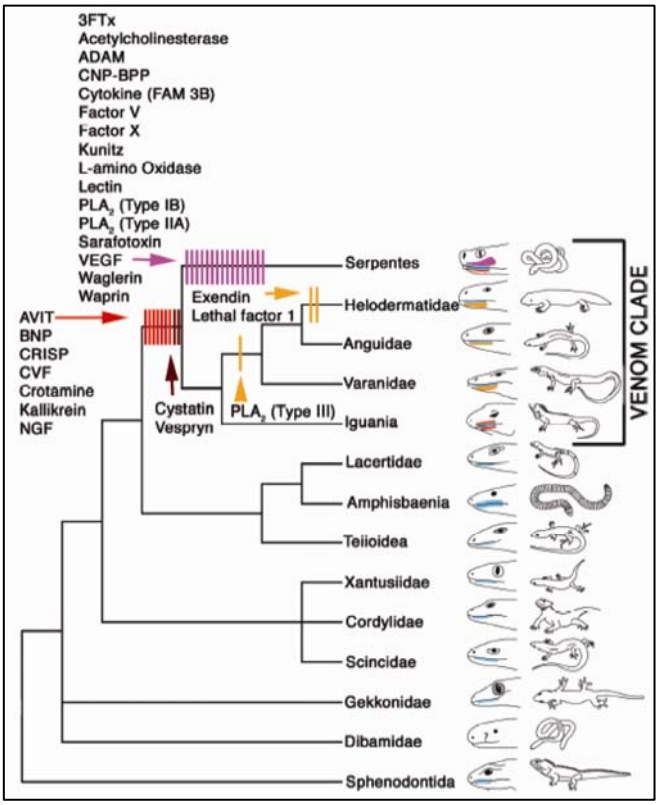


Figura 3. Detalle del árbol filogenético de los reptiles mostrando eventos sucesivos de reclutamiento de proteínas endógenas a lo largo de la divergencia evolutiva del clado de los reptiles escamosos venenosos (Fry et al. 2006). 3FTx, "*three finger toxin*"; ADAM, "*A Disintegrin And Metalloprotease*"; CNP-BPP, "*C-natriuretic bradykinin-potentiating peptide*"; VEGF, "*Vascular endothelial growth factor*"; CRISP, "*Cysteine-rich secretory protein*";

Las familias de proteínas presentes en los venenos de serpientes de la familia Viperidae incluyen enzimas (metalloproteasas dependientes de Zn^{2+}, fosfolipasas A_2 de tipo II, serinproteasas, L-amino acido oxidasa) y proteínas

sin actividad enzimática (péptidos natriuréticos, disintegrinas, inhibidores de proteasas tipo Kunitz, cistatina, lectinas específicas de galactosa, lectinas tipo C, factores de crecimiento vascular, CRISP). Sin embargo, el perfil proteico y la abundancia relativa de los diferentes grupos de proteínas varía ampliamente de especie a especie. La figura 4 ilustra este punto mostrando la composición de los venenos de las sepientes tunecinas *Cerastes cerastes cerastes*, *Cerastes vípera* y *Macrovipera lebetina transmediterranea* (Bazaa et al. 2005) y *Bitis gabonica gabonica* (Calvete et al. 2007).

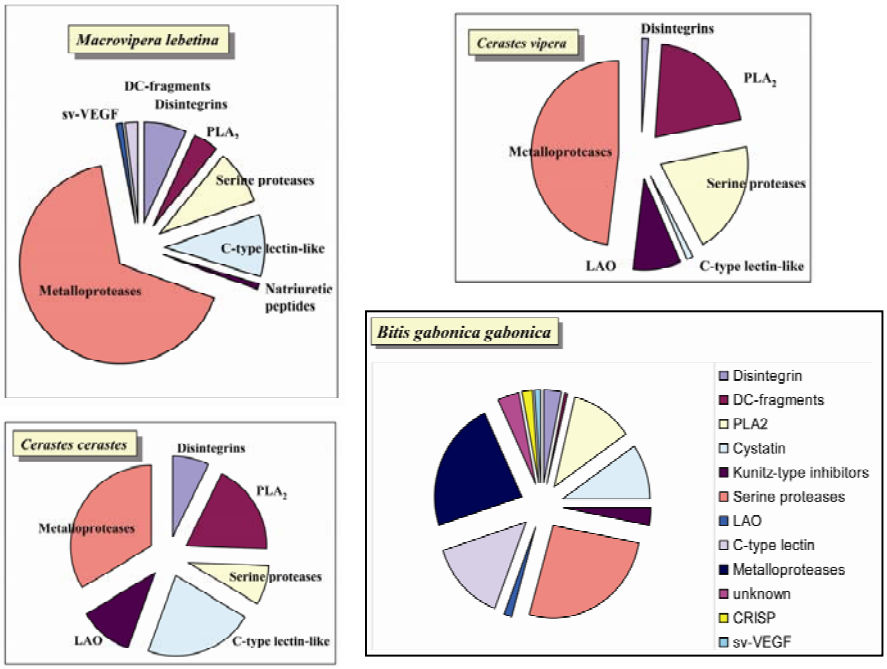


Figura 4. Distribución relativa de las familias de toxinas en diferentes serpientes de la familia Viperidae.

Es asimismo importante resaltar la gran variedad de isoformas de cada familia proteica presentes en todos los venenos analizados. Ello contribuye a la complejidad en composición proteica y a la gran diversidad de los efectos

biológicos de los venenos. La presencia de múltiples isoformas de una proteína evidencia la ocurrencia de duplicaciones génicas, y la gran diversidad estructural y funcional en el seno de cada familia multigénica indica la acción de una evolución acelerada de las toxinas (Kordis et al. 2002; Ohno et al. 2002). No es, pues, de extrañar que se hayan descrito multitud de actividades biológicas asociadas a miembros de la misma familia proteica. Así, diferentes isoenzimas de las fosfolipasas A_2 provocan hemolisis, miotoxicidad, neurotoxicidad, cardiotoxicidad, edemas y actividad anti- o procoagulante (Kini 2004; Kini 2005c; Ohno et al. 2003).

Las serinproteasas de venenos de serpientes interfieren con los mecanismos fisiológicos de coagulación, agregación plaquetaria, fibrinolisis y el sistema del complemento. Este grupo de enzimas incluye activadores de la proteína C y de los factores V, X, XI, protrombina, inactivadores de los factores Va y VIIIa, así como enzimas que proteolizan selectivamente el fibrinopéptido A o el B del fibrinógeno produciendo desfibrinogenemia y hemorragia (Kini 2005a; Kini 2005b; Kini 2006; Markland 1998).

Las metalloproteasas dependientes de Zn^{2+}, también denominadas hemorraginas, son las toxinas más abundantes de los venenos de víboras y serpientes de cascabel (Fig.4). Son endoproteasas que degradan proteínas de la matriz extracelular (colágeno, fibronectina, laminina, etc.) produciendo hemorragia local, inflamación y, a menudo, necrosis (Fox y Serrano 2005b; Gutierrez et al. 2005; Hati et al. 1999; Lu et al. 2005b). Esta familia incluye también α- o β-fibrinogenasas que degradan las regiones C-terminales de las cadenas A o B del fibrinógeno inhibiendo la formación de fibrina y la formación del coágulo (Swenson y Markland 2005). Otra metalloproteasa, jararhagina, del veneno de *Bothrops jararaca*, inhibe la agregación plaquetaria degradando al receptor plaquetario de colágeno, la integrina $\alpha_2\beta_1$ (Kamiguti et al. 1996).

Estructuralmente, las metaloproteasas de venenos de serpientes (SVMPs, Snake Venom MetalloProteases) se incluyen, junto con las ADAMs celulares (A Disintegrin And Metalloprotease) (White 2003) en la familia M12 (reprolisinas) de las metaloproteasas (Fox y Serrano 2005b). Tanto las SVMPs como las ADAM de las que derivaron por evolución divergente (Moura-da-Silva et al. 1996) son proteínas multidominio (Fig.5 y Fig. 6).

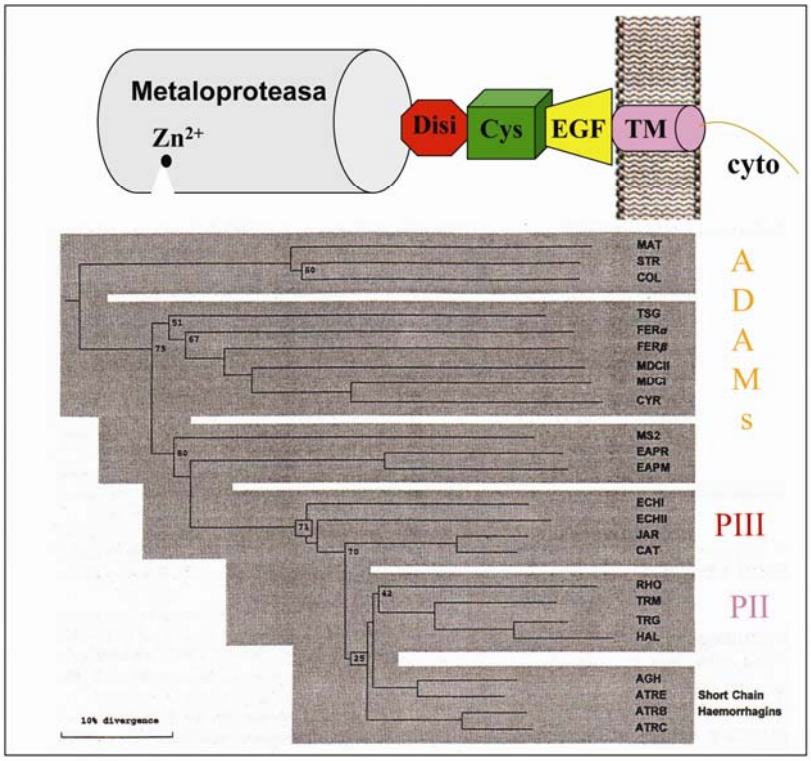


Figura 5. Esquema de la estructura modular de una metaloproteasa ADAM integral de membrana y modelo de evolución divergente de las SVMPs PIII y PII a partir de un precursor ADAM por pérdida de los dominios tipo EGF, transmembrana (TM) y citoplasmático (cyto).

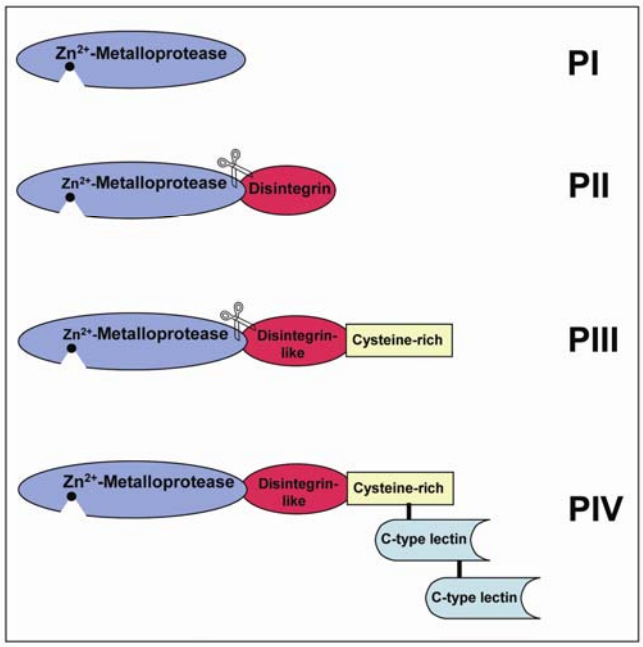


Figura 6. Clasificación y esquema de la estructura multidominio de las metaloproteasas dependientes de Zn^{2+} de veneno de serpientes de la familia Viperidae.

Las SVMPs de clase PI (20-30 kDa) contienen únicamente el dominio de metaloproteasa. Las SVMPs de clase PII (30-60 kDa) contienen además un dominio disintegrina C-terminal al dominio catalítico. Las SVMPs de clase PIII (60-90 kDa) son similares a las de clase II pero contienen un dominio C-terminal rico en cisteínas. Por último, las SVMPs PIV se diferencian de las PIII en que contienen dominios de lectina tipo C unidos mediante enlaces disulfuro entre si y al dominio rico en cisteínas (Fox y Serrano 2005b). Éstas últimas son relativamente raras como consecuencia de que la unión de un dominio de lectina tipo C al dominio rico en cisteínas es un proceso post-traduccional que requiere la existencia de una cisteína libre no conservada en las SVMPs PIII (Fox y Serrano 2005b).

Las metaloproteasas de clase PII (y algunas PIII) sufren procesamiento proteolítico liberando el dominio catalítico tipo PI y el dominio

de disintegrina (o el tándem de dominios tipo disintegrina y rico en cisteínas, también denominado fragmento DC) (Kini y Evans 1992). La actividad biológica de los fragmentos DC no está bien establecida. Por otra parte, como se describe más adelante con más detalle, las disintegrinas derivadas de SVMPs PII representan una grupo de antagonistas de receptores de la familia de las integrinas (Calvete 2005; Calvete et al. 2005). Por regla general, la actividad hemorrágica más potente se asocia a las SVMPs PIII (Gutierrez y Rucavado 2000). La reciente determinación de la estructura de la SVMP PIII VAP-1 del veneno de *Crotalus atrox* (Takeda et al. 2006) (Fig.7) dota de una base estructural a las observaciones bioquímicas. En efecto, la estructura cristalina muestra cómo el dominio rico en cisteínas adopta una conformación que sugiere que puede actuar de anclaje de la metaloproteasa, actuando sinérgicamente con el dominio catalítico en la degradación de la proteína sustrato.

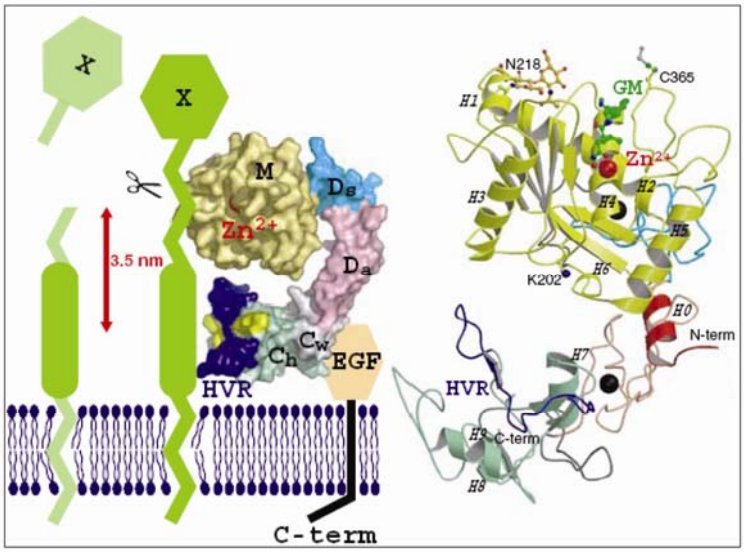


Figura 7. Estructura cristalina (derecha) y posible mecanismo de acción (izquierda) de la metaloproteasa PIII VAP1 mostrando la disposición espacial relativa de los dominios catalítico (M), tipo disintegrina (D) y rico en cisteínas (C).

Las toxinas que no presentan actividad enzimática, como las proteínas de la familia de las lectinas tipo C, también presentan diversidad de estructuras cuaternarias (dímeros $\alpha\beta$, dímeros de dímeros $(\alpha\beta)_2$, tetrámeros de dímeros $(\alpha\beta)_4$) (Fukuda et al. 2000; Horii et al. 2004; Mizuno et al. 1997; Murakami et al. 2003) y actividades biológicas pro- o anti- agregante de plaquetas utilizando para ello diversos mecanismos (activación del receptor de fibrinógeno -integrina $\alpha_{IIb}\beta_3$-, agonistas del receptor de colágeno -integrina $\alpha_2\beta_1$-, unión a los factores de coagulación IX y X, unión a las glicoproteínas GPVI o GPIb, interacción con el factor von Willebrand) (Clemetson et al. 2005; Lu et al. 2005a; Morita 2004; Morita 2005).

Las proteínas CRISP (Cysteine-Rich Secretory Proteins) están ampliamente distribuídas en los venenos de serpientes de las familias Viperidae y Elapidae de diferentes continentes (Yamazaki et al. 2003). Aunque sus propiedades funcionales son esencialmente desconocidas, el hecho de que la holothermina, una toxina de la familia CRISP del lagarto mexicano *Holoderma horridum horridum*, altere la función de diversos canales de Ca^{2+} y K^+ dependientes de voltaje y receptores de rianodina, sugiere que las CRISPs pueden representar una familia de toxinas que afectan a la contractilidad muscular. Esta hipótesis está sustentada por la observación de que el dominio rico en cisteínas de la estructura cristalina de la proteína Stecrisp del veneno de *Trimeresurus stejnegeri* presenta gran similitud estructural con las proteínas Bgk y Shk de la anémona marina *Bunodosoma granulifera* que bloquean canales de K^+ Kv1 sensibles a voltaje (Guo et al. 2005).

1.4 LA FAMILIA DE LAS DISINTEGRINAS

Las disintegrinas, objeto de esta Tesis, son polipéptidos (41-84 aminoácidos), cuya estructura está fuertemente tramada por 4-7 enlaces disulfuro, que son liberados al veneno de serpientes de la familia Viperidae por procesamiento proteolítico de SVMPs de clase PII y constituyen una familia de antagonistas de receptores de la familia de las integrinas (Calvete 2005; Calvete et al. 2005; McLane et al. 1998; McLane et al. 2004). Las primeras disintegrinas fueron descritas en el laboratorio de Stephan Niewiarowski (Temple University, Philadelphia, USA) a finales de la década de 1980 como potentes inhibidores de la agregación plaquetaria (Huang et al. 1987; Niewiarowski et al. 1994). Numerosos estudios bioquímicos y estructurales (revisados en el Trabajo 1 del Anexo) pusieron en evidencia que el mecanismo antiagregante de las disintegrinas se debía a su unión al receptor plaquetario de fibrinógeno, integrina $\alpha_{IIb}\beta_3$, utilizando para ello un tripéptido RGD (KGD o WGD) localizado en el ápice de un bucle móvil que sobresale 14-17 Å del cuerpo globular de la disintegrina (Monleon et al. 2005) y referencias citadas (Fig.7). Estas disintegrinas mimetizan el mecanismo de unión de la integrina a su ligando natural (Arnaout et al. 2005; Calvete 2004) como quedó elegantemente demostrado mediante la resolución de las estructuras cristalinas de los dominios extracelulares de la integrina $\alpha_v\beta_3$ en complejo con un péptido RGD cíclico y de la integrina $\alpha_{IIb}\beta_3$ (Calvete 2004; Del Gatto et al. 2006; Xiao et al. 2004; Xiong et al. 2002) (Fig.8).

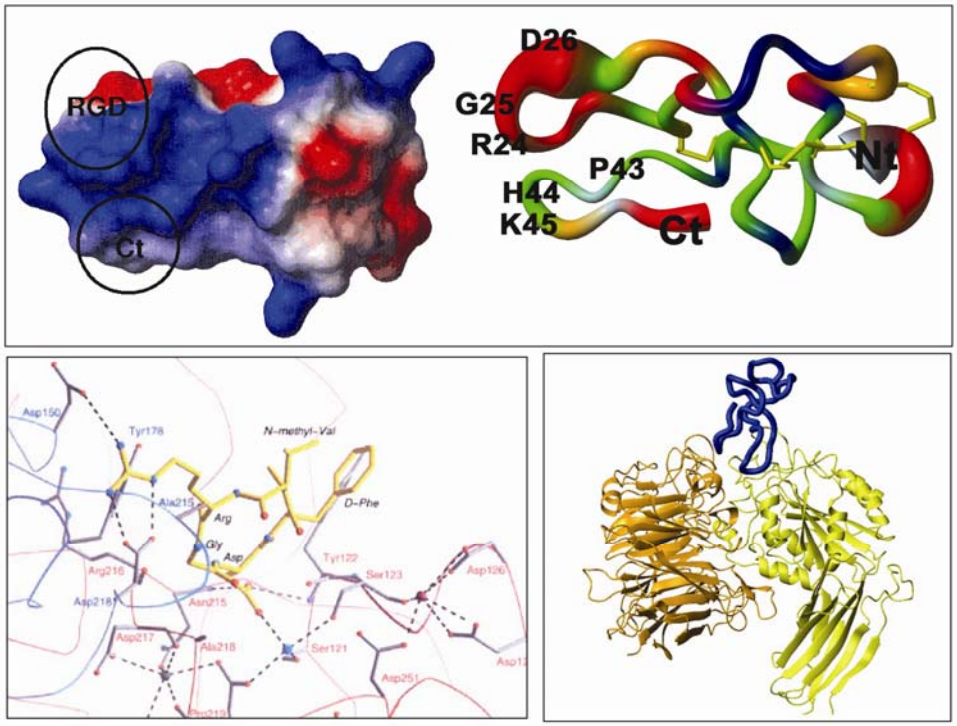


Figura 8. Estructura por RMN de la disintegrina echistatina (Monleon et al. 2005) mostrando la disposición relativa del tripéptido RGD del bucle de inhibición de integrinas y el extremo C-terminal. Estos dos elementos estructurales forman un epitopo funcional conformacional. Abajo, izquierda, esquema de las interacciones del péptido cíclico RGD en la estructura cristalina del complejo $\alpha_v\beta_3$-RGD (Xiong et al. 2002). El panel de la derecha muestra un modelo de la unión de echistatina (azul) al lugar de unión de ligandos de la integrina $\alpha_{IIb}\beta_3$ formado por elementos de ambas subunidades (α_{IIb}, amarillo; β_3, naranja).

Además del tripéptido mantenido en la conformación activa por enlaces disulfuro (Calvete et al. 1991), los aminoácidos adyacentes y el extremo C-terminal, que presenta movimiento concertado con el bucle de unión de integrinas (Monleon et al. 2005; Monleon et al. 2003) modulan la unión de la disintegrina a su receptor, induciendo cambios conformacionales en la integrina $\alpha_{IIb}\beta_3$ caracterizados por la expresión de epitopos LIBS (Ligand Induced Binding Sites) (McLane et al. 1998; Niewiarowski et al. 1994) cuya consecuencia es un aumento de la afinidad de la unión de la disintegrina a la

integrina $\alpha_{IIb}\beta_3$ y un incremento de su potencia inhibidora de la agregación plaquetaria (Marcinkiewicz et al. 1997).

Además de los motivos de inhibición plaquetaria (RGD, KGD, WGD), que también bloquean con diferente afinidad y potencia la unión de otras integrinas a sus ligandos naturales (ej. $\alpha_5\beta_1$ a fibronectina; $\alpha_8\beta_1$ a tenascina C, y $\alpha_v\beta_1$ y $\alpha_v\beta_3$ a vitronectina), se han caracterizado disintegrinas que poseen tripéptidos activos frente a otros sistemas de interacción integrina-ligando. Los motivos VGD y MGD presentan selectividad por la integrina $\alpha_5\beta_1$; el tripéptido MLD bloquea la función de las integrinas $\alpha_3\beta_1$, $\alpha_4\beta_1$, $\alpha_6\beta_1$, $\alpha_7\beta_1$, $\alpha_9\beta_1$ y $\alpha_4\beta_1$; y los motivos KTS y RTS antagonizan selectivamente la unión de colágeno I y IV a la integrina $\alpha_1\beta_1$ (Calvete 2005; Calvete et al. 2005) (Fig.9).

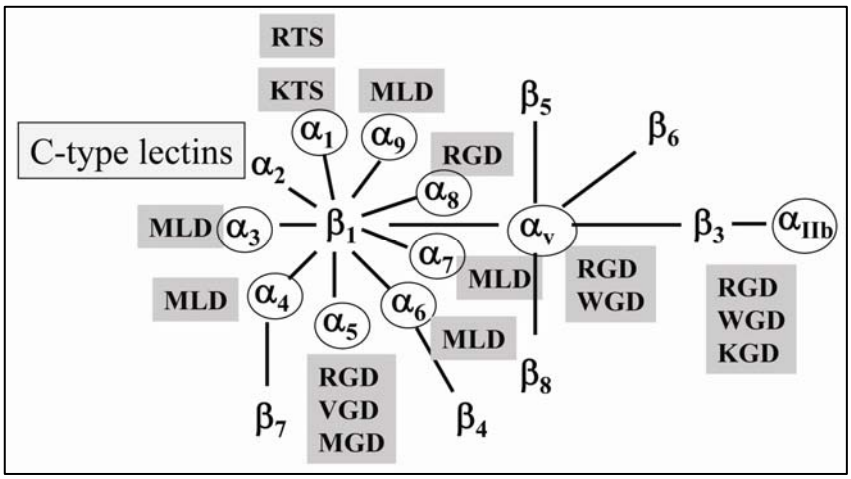


Figura 9. Esquema de la familia de las integrinas resaltando aquellos motivos estructurales de disintegrinas que bloquean su función.

Debe, pues, destacarse que la familia de las disintegrinas ha desarrollado un panel restringido, pero selectivo, de motivos de inhibición de

casi todas las integrinas de las familias β1 y β3. Las excepciones más notables son las integrinas linfocitarias ($\alpha_L\beta_2$, $\alpha_M\beta_2$ y $\alpha_X\beta_2$), que no son inhibidas por ninguna toxina conocida, y la integrina $\alpha_2\beta_1$, que -como se comentó anteriormente- es la diana de numerosas lectinas tipo C de veneno de serpientes. Además, el mapeo de los tripéptidos activos de disintegrinas sobre el árbol filogenético de las cadenas α de integrinas (que confieren la especificidad de unión de ligandos a heterodímeros que poseen una subunidad β común), sugiere una adaptación evolutiva de las disintegrinas a los lugares de unión de ligandos de las integrinas (Sanz et al. 2005) (Fig.10).

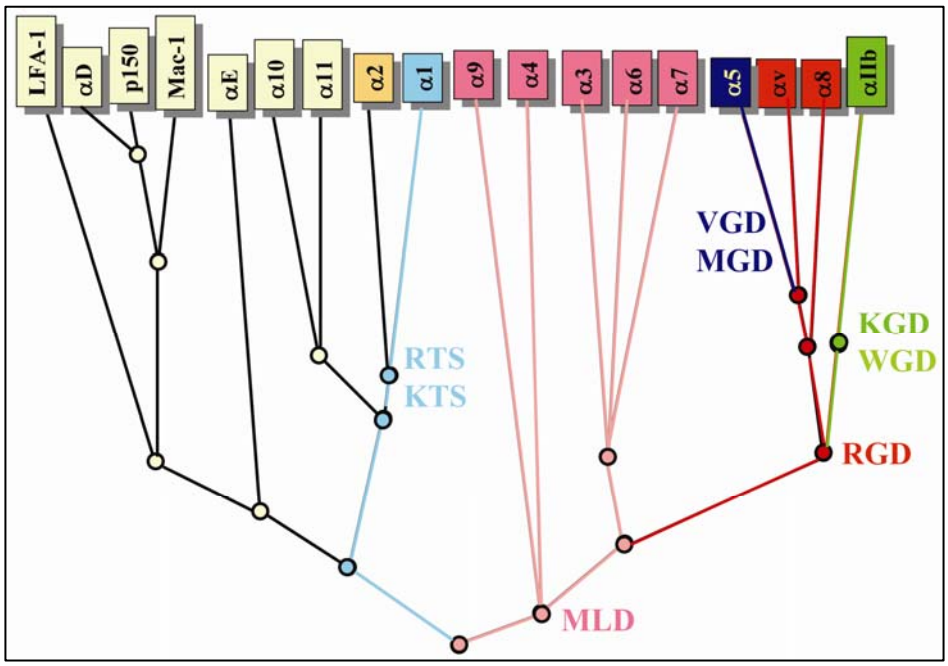


Figura 10. Arbol filogenético de las subunidades α de integrinas resaltando la adaptación evolutiva de los motivos estructurales de disintegrinas que bloquean su función.

Mientras que todas las disintegrinas largas conocidas y la inmensa mayoría de las disintegrinas medias y cortas descritas hasta la fecha expresan el tripéptido RGD, las disintegrinas diméricas presentan una mayor diversidad de motivos estructurales y de efectos biológicos (Marcinkiewicz 2005). Las integrinas diana de las diferentes disintegrinas participan en diversos procesos patológicos: la integrina $\alpha_{IIb}\beta_3$ es responsable de la formación de los agregados de plaquetas causantes de trombosis e isquemia cardíaca; la integrina $\alpha_v\beta_3$ desempeña un papel relevante en procesos de metástasis tumoral; las integrinas $\alpha_4\beta_1$, $\alpha_4\beta_7$ y $\alpha_9\beta_1$ participan en procesos de inflamación y autoinmunidad; las integrinas $\alpha_1\beta_1$ y $\alpha_v\beta_3$ han sido implicadas en el mecanismo de neovascularización (angiogénesis) tumoral. Los antagonistas de estas integrinas representan, por tanto, potenciales dianas terapéuticas (Curley et al. 1999; Gottschalk y Kessler 2002; Shimaoka y Springer 2003). A este respecto cabe destacar que las drogas Tirofiban (Aggrastat®) y Eptifibatide (Integrilin®) son inhibidores de la integrina $\alpha_{IIb}\beta_3$ basados en la secuencia RGD, de uso clínico para la prevención de episodios tromboembolíticos. En el Trabajo 2 del Anexo describimos la producción recombinante de la disintegrina jerdostatina de *Trimeresurus jerdonii*. Jerdostatina posee el tripéptido RTS y forma con obtustatina (*Vipera lebetina obtusa*), viperistatina (*Vipera palestinae*) y lebestatina (*Macrovipera lebetina transmediterranea*) el grupo de disintegrinas cortas que inhiben selectivamente a la integrina $\alpha_1\beta_1$ *in vitro* y la angiogénesis *in vivo* (Kisiel et al. 2004; Marcinkiewicz et al. 2003; Olfa et al. 2005).

La figura 11 muestra el mínimo número de mutaciones necesarias para convertir los diferentes tripéptidos entre sí. No obstante, la elucidación de los mecanismos evolutivos y la cadena temporal de eventos de esta diversificación funcional de las disintegrinas requiere un estudio detallado de los genes, inviable actualmente por la ausencia prácticamente total de información genómica.

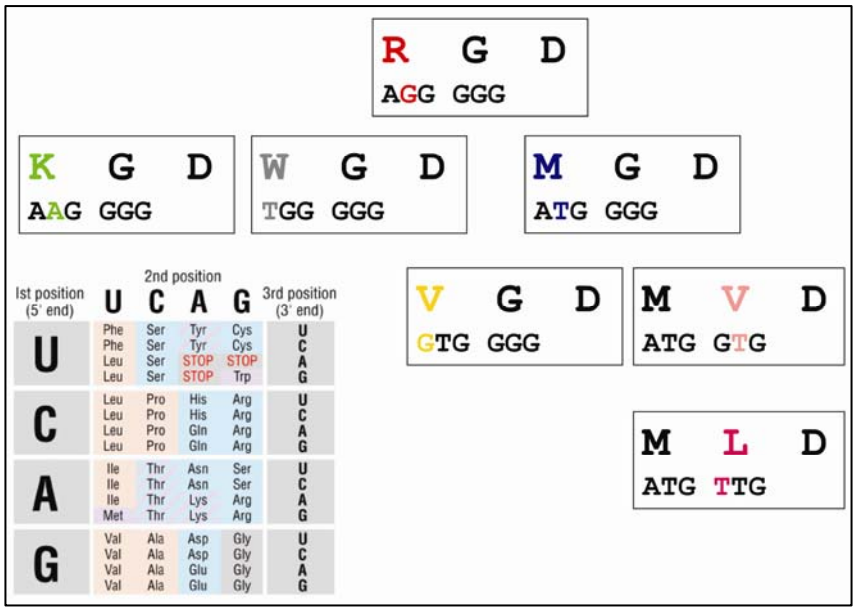
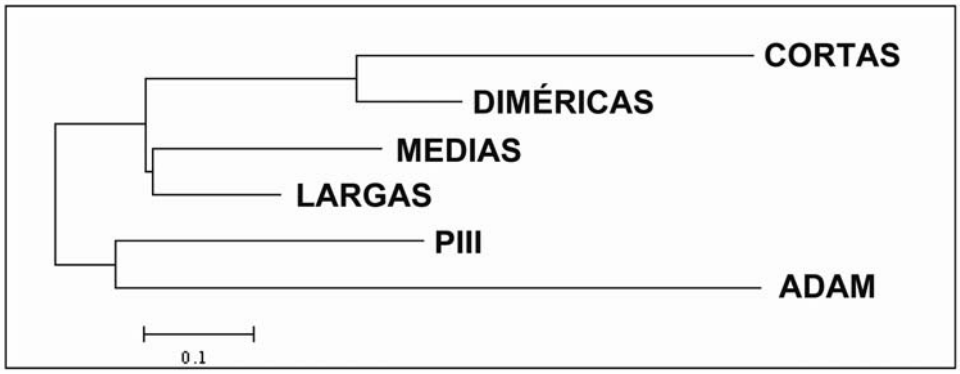


Figura 11. Mutaciones mínimas teóricas para interconvertir los diferentes motivos funcionales de las disintegrinas.

No obstante, lo que si que parece claro es que las diferentes clases de disintegrinas derivadas de SVMPs PII (Largas - ~84 aminoácidos y 7 puentes disulfuro (SS)-, Medias - ~70 resíduos y 6 SS -, Diméricas - homo- o heterodímeros de subunidades de ~63 aminoácidos incluyendo 4 SS por subunidad y 2 SS intercatenarios -, y Cortas - ~40-49 aminoácidos y 4 SS -) se originaron por duplicación génica y divergencia evolutiva mediante un mecanismo de minimización de la estructura primaria y pérdida sucesiva de enlaces disulfuro ("ingeniería de enlaces disulfuro") (Calvete et al. 2003) (Fig.12).



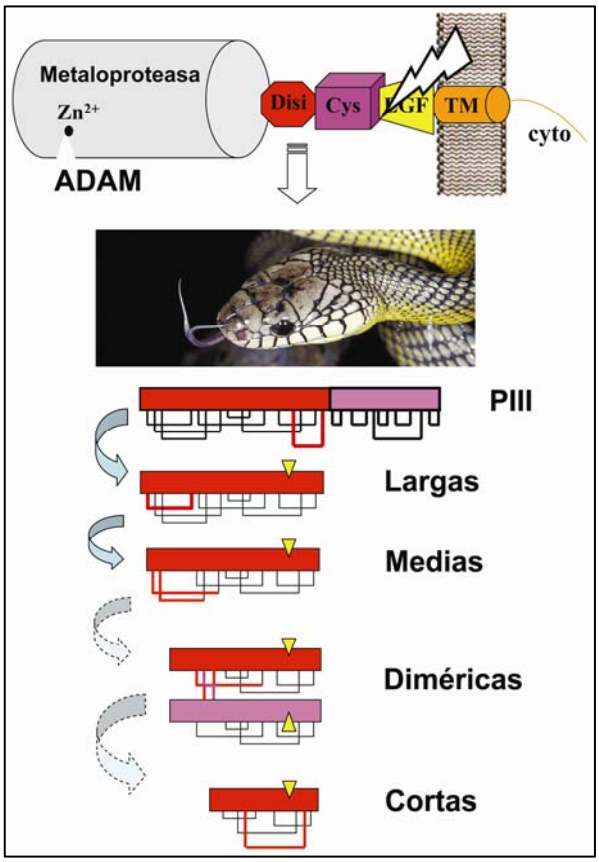


Figura 12. Panel superior, relación filogenética de los diferentes grupos de disintegrinas derivadas de SVMP PII, entre sí y con sus antecesoras SVMP PIII y ADAM. El panel de la izquierda muestra el esquema de diversificación estructural de las disintegrinas basado en el reclutamiento y diversificación estructural en la glándula del veneno de serpientes de los dominios extracelulares (metaloproteasa, tipo disintegrina y rico en cisteínas) de una proteína de la familia ADAM mediante un mecanismo de pérdida sucesiva de los enlaces disulfuro que se muestran en negrita y en rojo ("ingeniería de enlaces disulfuro") (Calvete et al. 2003). En las proteínas de venenos de serpientes, el dominio metaloproteasa se ha omitido para simplificar la figura. Los triángulos amarillos indican la posición de las secuencias de unión a integrinas.

Al comienzo de este trabajo de Tesis Doctoral este esquema de diversificación estructural postulaba la emergencia de las disintegrinas largas a partir de un precursor PIII por pérdida concertada del enlace disulfuro típico de estás últimas (entre las cisteínas XIII y XVI del dominio tipo disintegrina) (Calvete et al. 2000) y del dominio rico en cisteínas. Las causas de estos cambios estructurales eran la deleción del triplete TGT (pérdida de la CysXIII) y la aparición de un codón STOP en el dominio tipo disintegrina entre las cisteínas XV y XVI (Fig.13). No había ninguna evidencia experimental de la existencia de intermediarios en los que faltara únicamente uno solo de estos elementos estructurales (CysXIII-CysXVI o dominio rico en cisteínas).

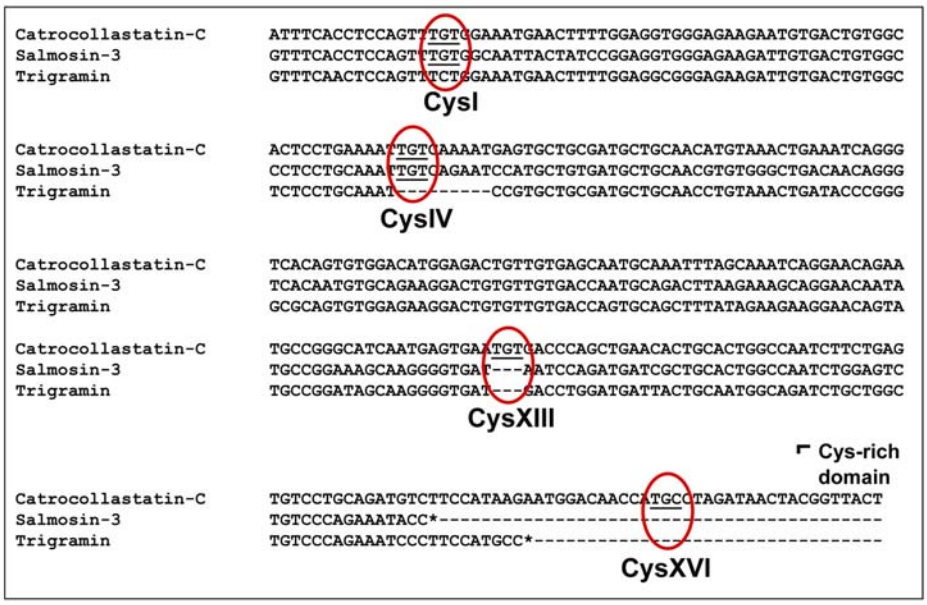


Figura 13. Comparación de las secuencias nucleotídicas del dominio tipo disintegrina de la SVMP PIII Catrocollastatina y las disintegrinas PII larga (Salmosina-3) y media (Trigramina), resaltando la pérdida de los pares de cisteínas XIII y XVI y del dominio rico en cisteínas entre Catrocollastatina y Salmosina-3, y del par CysI-CysIV entre Salmosina-3 y Trigramina.

Por otra parte, la comparación de secuencias de cDNAs de disintegrinas largas y medias indicaba que el paso de las primeras a las segundas implicaba la pérdida del par de cisteínas I y IV (Fig.12), las cuales en la disintegrina larga bitistatina forman un enlace disulfuro (Calvete et al. 1997). No se había clonado ningún DNA que codificara para subunidades de disintegrinas diméricas o para disintegrinas cortas y, por tanto, la ruta evolutiva para estas clases de disintegrinas admitía ambigüedades en el sentido de que tanto las disintegrinas diméricas como las cortas pudieran derivarse de las largas o de las medias. Con este trabajo de Tesis pretendimos encontrar respuestas a estos interrogantes.

2. OBJETIVOS

Objetivos

El objetivo general de esta Tesis Doctoral ha sido profundizar en el estudio de las bases moleculares de la diversificación estructural y funcional de la familia de las disintegrinas. Partimos de un esquema evolutivo basado en la pérdida sucesiva de enlaces disulfuro ("ingeniería de enlaces disulfuro") y la acumulación de mutaciones ("química combinatorial") en el bucle de inhibición de integrinas. Aplicamos técnicas proteómicas para determinar la composición proteica de venenos de serpientes con la doble perspectiva de estudiar la abundancia relativa de toxinas ("venómica", Artículo 1) y en particular de disintegrinas. Nos centramos posteriormente en el análisis de librerías de cDNA de las glándulas de Duvernoy (del veneno) de *Bitis arietans* (que expresa la disintegrina larga bitistatina) y de *Echis ocellatus* (cuyo veneno contiene las disintegrinas diméricas EO4 y EO5 y la disintegrina corta ocelatusina) con objeto de clonar mensajeros que pudieran representar intermediarios evolutivos de las familias de disintegrinas expresadas en los venenos. Habiendo identificado cDNAs que codifican a un intermediario (BA-5A) en la ruta evolutiva de la disintegrina bitistatina a partir del extremo C-terminal de una metaloproteasa dependiente de Zn^{2+} de clase PIII (Artículo 2), y cDNAs que codifican a precursores de la disintegrina ocelatusina a partir de una subunidad de disintegrina dimérica (Artículo 3), procedimos a investigar la organización genómica de genes de disintegrinas diméricas y cortas en *Echis ocellatus* y en *Macrovipera lebetina transmediterranea* (Artículo 4).

3. MATERIALES Y MÉTODOS. RESULTADOS

3.1. ARTÍCULO 1: Snake venomics: characterization of protein families in *Sistrurus barbouri* venom by cysteine mapping, N-terminal sequencing and tandem mass spectrometry analysis

Proteomics 2004, *4,* 327–338 DOI 10.1002/pmic.200300628

Paula Juárez
Libia Sanz
Juan J. Calvete

Instituto de Biomedicina
de Valencia, C.S.I.C.,
Valencia, Spain

Snake venomics: Characterization of protein families in *Sistrurus barbouri* venom by cysteine mapping, *N*-terminal sequencing, and tandem mass spectrometry analysis

The protein composition of the crude venom of *Sistrurus barbouri* was analyzed by two-dimensional sodium dodecyl sulfate polyacrylamide gel electrophoresis. Proteins were separated by reversed phase high-performance liquid chromatography and characterized by *N*-terminal sequence analysis. The molecular mass and number of cysteine residues of the purified proteins were determined by matrix-associated laser desorption/ionization-time of flight mass spectrometry. Selected protein bands were subjected to in-gel tryptic digestion and peptide mass fingerprinting. Analysis of the tandem mass spectrometry spectra of selected doubly-charged peptide ions was done by collision-induced dissociation in a quadrupole-linear ion trap instrument. Our results show that the venom proteome of the pigmy rattlesnake *S. barbouri* is composed of proteins belonging to a few protein families, which can be structurally characterized by their disulfide bond contents.

Keywords: Mass spectrometry / *N*-terminal sequencing / *Sistrurus barbouri* / Snake venom protein families
PRO 0628

1 Introduction

Snake venoms contain complex mixtures of hundreds of pharmacologically active molecules, including organic and mineral components (histamine and other allergens, polyamines, alkaloids), small peptides and proteins [1, 2]. The biological effects of venoms are complex because different components have distinct actions and may, in addition, act in concert with other venom molecules. The synergistic action of venom proteins may enhance their activities or contribute to the spreading of toxins. According to their major toxic effect in animals, snake venoms may conveniently be classified as neurotoxic or haemorrhagic. Among the first group, *Elapidae* snakes (mambas, cobras, and particularly the Australian snakes, which are well known to be the most toxic in the world) possess a wide variety of group I phospholipase A_2 (PLA_2; EC 3.1.1.4) isoenzymes. All known PLA_2 isozymes cleave the sn-2 acyl chain of glycerophospholipids to produce two potent lipid mediators: arachidonate, the key eicosanoid precursor for the production of thromboxanes, prostaglandins, and leukotrienes, and lyso-phosphatidylcholine, a chemoattractant for circulating monocytes. Snake venom PLA_2 isoenzymes exhibit an array of pharmacological activities, such as presynaptic and postsynaptic neurotoxicity, myotoxicity, cardiotoxicity, anticoagulant effects, platelet aggregation (inhibition or initiation), antihaemorrhagic activities, convulsant activities, hypotensive activities, oedema-inducing activities, and organ or tissue damage activities [2, 3]. The venom of *Elapidae* snakes is also a rich source of three-fingered neurotoxins, 60–70 amino acid polypeptides whose three-dimensional structures are highly conserved, but which exert a wide range of activities on particular subtypes of voltage-dependent ion channels from the brain and neuromuscular junctions [4].

Venoms of *Viperidae* and *Crotalidae* snakes (vipers and rattlesnakes) contain a number of different proteins that interfere with the coagulation cascade, the normal haemostatic system and tissue repair. Consequently, envenomenations by these snakes generally results in persistent bleeding. These proteins can be grouped into a few major protein families, including enzymes (serine proteinases, Zn^{2+}-dependent metalloproteases of the reprolysin family, and group II PLA_2 isoenzymes) and proteins with no enzymatic activity (C-type lectins, and disintegrins) [3].

Serine proteinases are thrombin-like enzymes, which trigger the clotting of fibrinogen thereby inducing platelet aggregation. Snake venom metalloproteases induce local

Correspondence: Dr. Juan J. Calvete, Instituto de Biomedicina de Valencia, C.S.I.C., c/Jaime Roig, 11, 46010 Valencia, Spain
E-mail: jcalvete@ibv.csic.es
Fax: +34-96-369-0800

Abbreviation: PLA_2, phospholipase A_2

hemorrhaging as a primary consequence of degradation of extracellular matrix proteins, while PLA_2 causes severe local swelling followed by necrosis. C-type lectin-like proteins are multimeric molecules [5–7], which include inhibitors and activators of coagulation factors V (AaACP), IX, and X (botrocetin, fIX/X-binding protein); proteins that bind to the platelet membrane GPIb/IX complex inhibiting von Willebrand factor binding to this receptor and either block (echicetin, agkicetin, flavocetin, and tokarecetin) or promote (alboaggregins A and B) platelet aggregation; and potent activators (colvuxin) of the platelet collagen receptor GPVI, and selective inhibitors (EMS16, rhodocetin) of the platelet integrin $\alpha 2\beta 1$ [3]. Disintegrins are released from venoms by proteolytic processing of PII Zn^{2+}-metalloproteinases, and inhibit integrin-ligand interactions [3, 8]. NMR studies of several short (echistatin) and medium (kistrin, flavoridin, albolabrin) disintegrins revealed that the active tripeptide is located at the apex of a mobile loop protruding 14–17 Å from the protein core [9–12]. RGD-containing disintegrins show different binding affinity and selectivity towards integrins which recognize the RGD sequence in their ligands (*i.e.* $\alpha_{IIb}\beta_3$, $\alpha_v\beta_3$ and $\alpha_5\beta_1$) [8]. KGD-containing barbourin inhibits $\alpha_{IIb}\beta_3$ integrin with a high degree of selectivity [13]. The MLD sequence is responsible for the inhibitory activity of EC3, VLO5 and EO5 towards the α_4 integrins [14, 15]. Selective recognition of $\alpha_5\beta_1$ by EMF-10 is associated with the MGD(W) sequence [16]. The presence of a WGD motif in CC8, a heterodimeric disintegrin isolated from the venom of the North African sand viper, *Cerastes cerastes cerastes*, has been reported to increase its inhibitory effect on $\alpha_{IIb}\beta_3$, $\alpha_v b_3$and $\alpha_5\beta_1$ integrins [17].

Toxic venom proteins play a number of adaptatives roles: immobilizing, paralyzing, killing, liquefying prey and deterring competitors. It is assumed that the existence in the same venom of a diversity of proteins belonging to the same family but which differ from each other in their pharmacological effects (neurotoxins [3], PLA_2 [3], C-type lectin-like molecules [18], metalloproteases [19], disintegrins [15], serine proteinases [20]) reflects an accelerated Darwinian evolution. Hence, gene duplication creates redundancy and allows a gene copy to escape the pressure of negative selection and evolve a new function. The fact that members of a single family show remarkable structure similarity but differ in their biological targeting makes them valuable biotechnological tools for studying physiological processes and provides exiting challenges for delineating structure-function correlations.

Neurotoxins have been used to map functional epitopes on nicotinic and muscarinic acetylcholine receptors, and on potassium and calcium channels, and represent potential lead compounds for designing selective ligands for a molecular isoform of an ion channel or receptor family. Snake venom components affecting thrombosis and haemostasis are useful tools for investigating blood coagulation mechanisms and have been extensively used in the development of diagnostic tests. Disintegrins are valuable tools for identifying novel integrin-binding sequence motifs which may shed light on the structural requirements of selective integrin inhibition. On relatively rare occasions, toxins themselves have been used as therapeutic agents [21]. Disintegrins have found numerous applications in studies on platelet thrombosis, angiogenesis, cancer, bone destruction, and inflammation. They have been used to prevent experimental arterial thrombosis in animal models, and a synthetic derivative of the disintegrin barbourin, integrilin, has been developed as a drug that prevents arterial thrombosis after angioplasty [3, 8]. It is also noteworthy that the natural resistance to envenomation by snakes observed in a few warm-blooded animals as well as in several snakes can be explained, in most cases, by the presence of proteins in the sera from resistant animals which can be grouped as inhibitors of either PLA_2 (antimyotoxic and antineurotoxic factors) or metalloproteinases of the reprolysin family (antihemorrhagic factors) [3, 22, 23]. Thus, establishing structure-function relationships of isolated venom toxins may lead to the design of novel nontoxic drugs for clinical use in cases of severe envenomations. The aim of this study was to analyze the protein composition of snake venoms with the perspective of isolating and characterizing novel proteins for structural and functional investigations. Here we report the proteomic analysis of the venom of pygmy rattlesnake *Sistrurus barbouri*.

2 Materials and methods

2.1 2-D SDS-PAGE

Lyophilized venom from *Sistrurus barbouri* was purchased from Latoxan Serpentarium (Rosans, France). Its protein composition was analyzed by 2-D using an IPGphor (Amersham Bioscience, Uppsala, Sweden) instrument. For the first dimension (IEF), 500–1000 μg of total venom proteins (in 250 μL of 8 M urea, 4% CHAPS and 0.5% IPG buffer) were loaded on a 13 cm IPG strip (pH range 3–10) using the following focusing conditions: 30 V for 6 h, 60 V for 6 h, 500 V for 1 h, 1000 V for 1 h, and 8000 V for 2 h. Electrophoretic separation (second dimension) was done in a 16 cm 15% acrylamide gel. CBB was employed for protein staining.

2.2 Isolation of proteins

For RP HPLC separations, 2–5 mg of the crude venom was dissolved in 100 μL of 0.05% TFA and 5% ACN, and insoluble material was removed by centrifugation in an Eppendorf centrifuge (Hamburg, Germany) at 13 000 × *g* for 10 min at room temperature. Proteins in the soluble material were separated using an ETTAN™ LC HPLC system (Amersham Biosciences) and a Lichrospher RP100 C18 column (250 × 4 mm, 5 μm particle size; Merck, Darmstadt, Germany) eluted at 1 mL/min with a linear gradient of 0.1% TFA in water (solution A) and ACN (solution B) (isocratically (5% B) for 5 min, followed by 5–45% B for 90 min, and 45–70% B for 20 min). Protein detection was at 215 nm. Peaks were collected manually and dried in a Speed-Vac (Savant, Holbrook, NY, USA).

2.3 Characterization of isolated proteins

Isolated proteins (2–5 mg/mL in 100 mM ammonium bicarbonate, pH 8.3, containing 5 M guanidinium hydrochloride) were reduced with 1% v/v 2-mercaptoethanol for 2 min at 85°C, alkylated by addition of 4-vinylpyridine (5% v/v final concentration) and incubated for 1 h at room temperature. The *S*-pyridylethylated proteins were subjected to *N*-terminal sequence analysis using a Procise instrument from Applied Biosystems (Foster City, CA, USA) following the manufacturer's instructions. Amino acid sequence similarity searches were performed against available databanks using the BLAST program [24] implemented in the WU-BLAST2 search engine at http://www.bork.embl-heidelberg.de. The molecular masses of the purified proteins were determined by MALDI-TOF MS using an Applied Biosystems Voyager-DE Pro mass spectrometer operated in linear mode. To this end, equal volumes (0.5 μL) of the protein solution and the matrix (sinapinic acid; Sigma, St. Louis, MO, USA), saturated in 50% ACN and 0.1% TFA, were mixed onto the MALDI-TOF plate. The mass calibration standard consisted of a mixture of the following proteins, whose isotope-averaged molecular mass in Daltons are given in brackets: bovine insulin (5734.6), *Escherichia coli* thioredoxin (11674.5), horse apomyoglobin (16 952.6), *E. coli N*-acetyl-*L*-glutamate kinase (NAGK; 27 159.5) [25], *Pyrococcus furiosus* Carbamoyl-phosphate synthetase (PFU; 34 297.4) [26], *Parkia platycephala* seed lectin (PPL; 47 946) [27], and bovine serum albumin (66 431). NAGK and PFU were generous gifts from Dr. Vicente Rubio (Instituto de Biomedicina de Valencia, Valencia, Spain). PPL was a generous gift from Dr. Benildo S. Cavada (Universidade Federal de Ceará, Fortaleza, Brazil). The other proteins were purchased from Applied Biosystems.

Isolated proteins were lyophilized, resuspended at a concentration of 2–5 mg/mL in 100 mM ammonium bicarbonate, pH 8.3, and degraded with trypsin (1:100 w/w, enzyme to substrate ratio) for 18 h at 37°C. The tryptic peptide mixture was lyophilized, dissolved in 0.1% TFA, and 0.85 μL was spotted onto the stainless steel sample plate of a MALDI-TOF Voyager-DE Pro (Applied Biosystems) mass spectrometer. The sample was mixed on the plate with the same volume of a saturated solution of α-cyano-4-hydroxycinnamic acid (Sigma) in 50% ACN containing 0.1% TFA, dried, and analyzed in delayed extraction and reflectror modes. A tryptic peptide mixture of *Cratylia floribunda* seed lectin (SWISS-PROT accession number P81517) prepared and previously characterized in our laboratory was used as mass calibration standard (mass range, 450–3300 Da).

2.4 Quantitation of free cysteine residues and disulfide bonds

For quantitation of free cysteine residues and disulfide bonds, the purified proteins were dissolved to a concentration of 2–5 mg/mL in 10 μL of 50 mM HEPES, pH 9.0, 5 M guanidinine hydrochloride, and 1 mM EDTA. They were heat-denatured at 85°C for 15 min, allowed to cool at room temperature, and incubated with either 10 mM 4-vinylpyridine [28, 29] for 1 h at room temperature, or with 10 mM 1,4-dithioerythritol (Sigma) for 15 min at 80°C. 4-Vinylpyridine, final concentration 25 mM, was added and samples were incubated for 1 h at room temperature. Pyridylethylated (PE) proteins were freed from reagents using a C18 Zip-Tip pipette tip (Millipore, Bedford, MA, USA) after activation with 70% ACN and equilibration in 0.1% TFA. Following protein adsorption and washing with 0.1% TFA, the PE proteins were eluted onto the MALDI-TOF plate with 1 μL of 70% ACN and 0.1% TFA and subjected to mass spectrometric analysis as described in Section 2.3.

The number of free cysteine residues (N_{SH}) was determined using Eq. 1:

$$N_{SH} = (M_{VP} - M_{NAT})/105.3 \quad \text{(Eq. 1)},$$

where M_{VP} is the mass of the denatured but nonreduced protein incubated in the presence of 4-vinylpyridine; M_{NAT} is the mass of the native, HPLC-isolated protein; and 105.3 is the mass increment due to the pyridylethylation of one thiol group.

The number of total cysteine residues (N_{Cys}) can be calculated from Eq. 2:

$$N_{Cys} = [(M_{PE} - M_{VP})/106.3] + N_{SH} \quad \text{(Eq. 2)},$$

where M_{PE} is the mass (in Da) of the reduced and pyridylethylated protein; and 106.3 is the mass increment due to the pyridylethylation of a cysteine residue, which, prior to reduction, was involved in the formation of a disulfide bond.

Finally, the number of disulfide bonds $N_{S\text{-}S}$ can be calculated from Eq. 3:

$N_{S\text{-}S} = (N_{cys} - N_{SH})/2$ (Eq. 3).

All mass values in Eqs. 1–3 are in Da. These equations are valid for single chain polypeptides.

2.5 In-gel enzymatic digestion and mass fingerprinting

Protein bands of interest were excised from a CBB-stained SDS polyacrylamide gel and subjected to automated reduction and alkylation with iodoacetamide, and digestion with sequencing grade bovine pancreas trypsin (Roche, Barcelona, Spain) using a ProGest digestor (Genomic Solutions, Chelmsford, MA, USA) following the manufacturer's instructions. The tryptic peptide mixtures were dried in a SpeedVac (Savant, Holbrook, NY, USA) and dissolved in 5 μL of 50% ACN and 0.1% TFA. Digests (0.85 μL) were spotted onto a MALDI-TOF sample holder, mixed with an equal volume of a saturated solution of α-cyano-4-hydroxycinnamic acid (Sigma) in 50% ACN containing 0.1% TFA, dried, and analyzed with an Applied Biosystems Voyager-DE Pro MALDI-TOF mass spectrometer, operated in delayed extraction and reflectror modes. The peptide mass fingerprint obtained for each spot was compared with the known trypsin digest protein nonredundant databases (releases of February 2003) from SWISS-PROT (http://us.expasy.org) or NCBI (http://www.ncbi. nlm.nih.gov) using the MS-Fit search engine of the Protein Prospector program (v.3.4.1) developed by the University of California at San Francisco and available at http:// prospector.ucsf.edu. All searches were constrained to a mass tolerance of 120 ppm.

2.6 CID MS/MS

For structure assignment confirmation or peptide sequencing, the protein digest mixture was loaded in a nanospray capilar and subjected to electrospray ionization mass spectrometric analysis using a QTrap mass spectrometer (Applied Biosystems) [30] equipped with a nanospray source (Protana, Odense, Denmark). Doubly- or triply-charged ions selected after enhanced resolution MS analysis were fragmented using the enhanced production with Q_0 trapping option. Enhanced resolution was performed at 250 amu/s across the entire mass range, a scanning mode that enables mass accuracy of less than 20 ppm making charge state identification reliable up to charge state 5. Enhanced production refers to the performance of the PE-SCIEX developed and patented LINAC™ (Q2) collision [30] cell technology, which accelerates ions through the collision cell thereby correcting the slow movement of ions due to high pressures existing within the chamber, and provides high sensitivity and improved resolution in MS/MS mode in comparison to triple quadrupoles without the LINAC™ collision cell. For MS/MS experiments, Q1 was operated at unit resolution, the Q1 to Q2 collision energy was set to 35 eV, the Q3 entry barrier was 8 V, the linear ion trap Q3 fill time was 250 ms, and the scan rate in Q3 was 1000 amu/s. CID spectra were interpreted manually or using the on-line form of the MASCOT program at http://www.matrixscience.com.

3 Results and discussion

3.1 Protein composition

Separation of the protein components of the crude venom of *S. barbouri* by 2-DE revealed the presence of three major protein spots (Fig. 1). Two of these proteins exhibited acidic p*I*s and molecular masses of 46 kDa and 14 kDa, while the other major protein had a basic p*I* and an apparent molecular mass of 12 kDa. A number of other protein bands migrated between the neutral and acidic part of the p*I* gradient and had molecular masses around 30 kDa or below 10 kDa (Fig. 1). To characterize these proteins, the crude venom was fractionated by RP HPLC. The proteins in the major RP HPLC-isolated peaks (Fig. 2) were identified by a combination of *N*-terminal se-

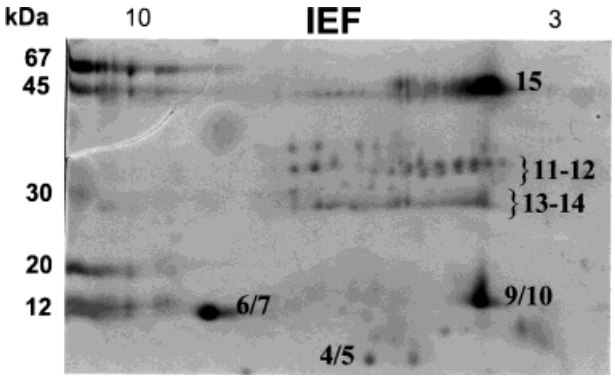


Figure 1. Two-dimensional gel electrophoresis of *S. barbouri* venom proteins. Total venom proteins (1000 μg) were subjected to IEF using a 13 cm IPG strip followed by SDS-PAGE in a 15% acrylamide (16 cm) gel. Spots assigned by peptide mapping and/or *N*-terminal sequencing are labelled and listed in Table 1.

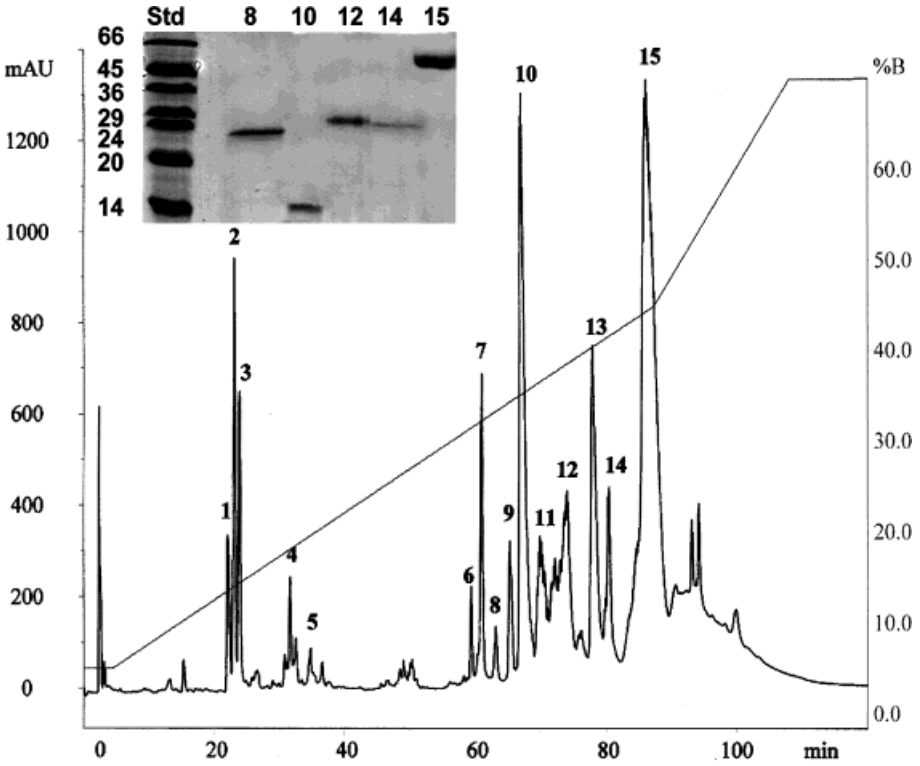


Figure 2. Separation, by RP HPLC, of *S. barbouri* venom proteins. *N*-terminal sequencing and MALDI-TOF mass spectrometric characterization of the major protein fractions are shown in Table 1. The insert shows an SDS-PAGE of purified proteins.

quencing and mass spectrometric determination of molecular masses and cysteine content (Table 1). In addition, comparison of their tryptic peptide mass fingerprints allowed us the tentative correlation of a number of RP HPLC-isolated proteins with 2-D electrophoretic spots (Fig. 1). However, with the exception of the protein in peak 15, which is described below, the peptide mass fingerprinting approach alone was unable to identify any proteins in the databases. This may reflect the almost complete absence of *S. barbouri* protein entries in the SWISS-PROT/TrEMBL and NCBI nonredundant databases, and the lack in the *S. barbouri* proteins of a significant set of tryptic peptides with identical masses in homologue proteins from other snake species represented in the databanks.

Fractions 1–3 contained low molecular mass peptides which were undetectable by SDS-PAGE. They were not further characterized. Except for the protein in peak 15, which had a blocked *N*-terminus, all other proteins could be assigned to known protein families (Table 1). Thus, the polypeptides recovered in RP HPLC fractions 4 and 5 clearly identified them as members of the medium-sized group (about 70 residues including 12 cysteines involved in 6 S-S bonds) of the disintegrin family (Table 1). Disintegrins are potent inhibitors of integrin receptors of the β1 and β3 families [8, 15]. In particular, the mass of the major polypeptide in fraction 4 (7500.7 Da, Table 1), matches accurately the mass of barbourin (7502.3 Da for residues 2–72 of the SWISS-PROT entry P22827), a previously reported integrin αIIbβ3-specific disintegrin [13]. Whether the other disintegrin molecules present in fractions 4 and 5 represent *N*- and/or *C*-terminal processed barbourin molecules or novel disintegrins requires further structural characterization.

Fractions 6, 7, 9 and 10 contained PLA_2 enzymes, which are widely expressed in snake venoms. They are characterized by the presence in their primary structures of 14 cysteine residues involved in the formation of seven disulfide bonds. This feature appears also to be conserved in *S. barbouri* PLA_2 enzymes (Table 1). Figure 3 illustrates the procedure followed for the quantitation of the number of cysteine residues in PLA_2 isolated in RP fraction 10 of Fig. 2. The closest homologue of the PLA_2s in fractions 6 and 7 is the acidic PLA_2 from *Protobothrops mucrosquamatus* (Q90W39) (96% identity), whereas the *S. barbouri* venom proteins in fractions 9 and 10 displayed highest sequence similarity (82% sequence identity) to PLA_2s from *Crotalus viridis viridis* (AA093140), *Crotalus atrox* (P00624), and *Agkistrodon piscivorus piscivorus* (A53872). The *N*-terminal sequences of the proteins recovered in fractions 11 and 12 identified them as serine proteinases closely related to enzymes found in venoms of *Viperidae* snakes exhibiting a variety of peptidase activities.

Table 1. *N*-terminal sequencing and MALDI-TOF mass spectrometric characterization of the major protein fractions purified by RP HPLC from the crude venom of the pigmy rattlesnake *S. barbouri* (Fig. 2).

HPLC fraction	*N*-terminal sequence	MALDI-TOF MS (Da)			No. of cysteines[d]			Protein family
		M_{Nat}[a]	M_{VP}[b]	M_{PE}[c]	Free SH	Total Cys	S-S	
4	AGEECDCGSP GEECDCGSPE EECDCGSPEN	7 501	7 502	8 762	–	12	6	Disintegrin
5	GEECDCGSPE EECDCGSPEN ECDCGSPENP	7 110	7 110	8 371	–	12	6	Disintegrin
6	NLLQFNKMIKIMT	13 952	13 953	15 442	–	14	7	PLA_2
7	NLLQFNKMIKIMTKKNAIP	13 956	13 956	15 455	–	14	7	PLA_2
8	SVNFDSESPPKPEIQ	24 841	24 840	26 506	–	16	8	CRISP
9	NLITFEQLIM	13 963	13 961	15 494	–	14	7	PLA_2
10	HLITFEQLIMKIAGRSGVFW	13 980	13 983	15 483	–	14	7	PLA_2
11	VIGGNECNINEHRSL	27 418	27 420	28 665	–	12	6	Ser-proteinase
12	VIGGDECNINEHRFL	27 430	27 431	28 765	–	12	6	Ser-proteinase
13	NPEHQRYVELFIVVDHGM	23 187	23 293	23 921	1	7	3	Metalloproteinase
14	NPEHQRYVELFIVVD	23 356	23 375	24 089	1	7	3	Metalloproteinase
15	Blocked	48 555	48 664	52 241	1	35	17	ADAM

a) M_{NAT}, mass of the native, HPLC-isolated protein
b) M_{VP}, mass of the denatured but nonreduced protein incubated in the presence of 4-vinylpyridine
c) M_{PE} ,mass (in Da) of the reduced and pyridylethylated protein
d) Quantitation of free cysteine residues (SH), total cysteine residues (Total Cys), and disulfide bonds (S-S) was done using Eqs. 1–3 as described in Section 2.4

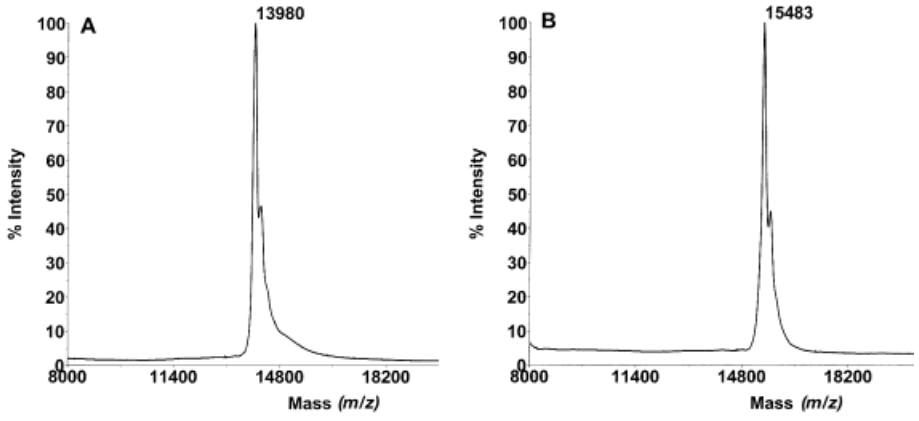


Figure 3. Quantitation of cysteine residues. A) MALDI-TOF mass spectrum of the native protein isolated in peak 10 of the RP HPLC separated *S. barbouri* venom proteins shown in Fig. 2. B) MALDI-TOF mass spectrum of the same protein as in A) after reduction and *S*-pyridylethylation. The number of cysteine residues was derived using Eq. 2 N_{Cys} = (15 483–13 980)/106.3 = 14.1, as described in Section 2.4.

3.2 Zn^{2+}-metalloproteases

The molecular mass (48.5 kDa) and presence of 35 cysteine residues *per* molecule (a free cysteine and 17 disulfide bonds; Table 1) strongly suggested that the *N*-terminal blocked protein might be a PIII metalloprotease, which are widely distributed in *Viperidae* and *Crotalidae* snakes. PIII metalloproteases (ADAMs or reprolysins) are mosaic proteins composed of *N*-terminal Zn^{2+}-metalloprotease (23 kDa, 7 cysteines) followed by a disintegrin domain (13 kDa, 16 cysteines) and a cysteine-rich domain (12 kDa, 12 cysteines), which exhibit the most potent

extracellular matrix degrading activity among hemorrhagic metalloproteases [31]. To confirm the identity of the *N*-terminal blocked protein, the protein band was excised from a CBB-stained SDS polyacrylamide gel and subjected to automated in-gel tryptic digestion. The poor quality of the MALDI-MS spectrum might be due to resistance to proteolysis of the 2-DE separated protein. Indeed, the carbamidomethylated protein-15 precipitated in the buffer used for tryptic digestion unless guanidinium hydrochloride (up to 1.5 M) was added to the solution. Nevertheless, ions at *m/z* 1052.30, 1230.37, 1285.48, 1552.56, and 2208.04 of the peptide mass fingerprint of the 48.5 kDa protein shown in Fig. 4 matched the polypeptide stretches ^{494}GNYYGYCR^{501}, ^{524}DNSPGQNNPCK^{534}, ^{535}MFYSNDDEHK^{544}, ^{558}VCSNGHCVDVATAY^{571}, and ^{199}YLYMHVALVGLEIWSNGDK^{217}, respectively, of the metalloproteinase jararhagin precursor from *Bothrops jararaca* venom (SWISS-PROT accession code P30431). This entry displayed rank number 1 and had a MOWSE score of 2.43×10^4. The second rank entry had a MOWSE score of 37. As a whole, the peptides cover 10.8% of the full-length 571-residue protein. The 198–217 peptide belongs to the Zn^{2+}-metalloproteinase domain, while the other peptides are all located within the cysteine-rich domain of jararhagin (Fig. 4).

Interestingly, peptide 558–571 represents the *C*-terminal fragment of jararhagin. As a whole, these peptides strongly suggest that the 48.5 kDa protein of *S. barbouri* venom may correspond to a full-length reprolysin Zn^{2+}-metalloprotease (also termed PIII SVMP). The structural assignment of the jararhagin-like tryptic peptides was confirmed by CID analysis of the doubly-charged ions at *m/z* 526.7 (Fig. 5), 615.9, and 776.1 (Fig. 6). The fragmentation pattern of the ion at $(M+2H)^{2+}$ 526.7 was unusual because most of the theoretical *b* and *y* ions have the same mass and are therefore indistinguishable in the mass spectra. Nevertheless, the data identified *m/z* 526.7 (2+) as GNYYGYCR, confirming the MALDI-TOF mass fingerprinting assignment.

Fragmentation of the doubly-charged ion at *m/z* 776.1 was induced at low Q1 to Q2 collision energy (35 eV) to avoid CID of the singly-charged ions of high molecular

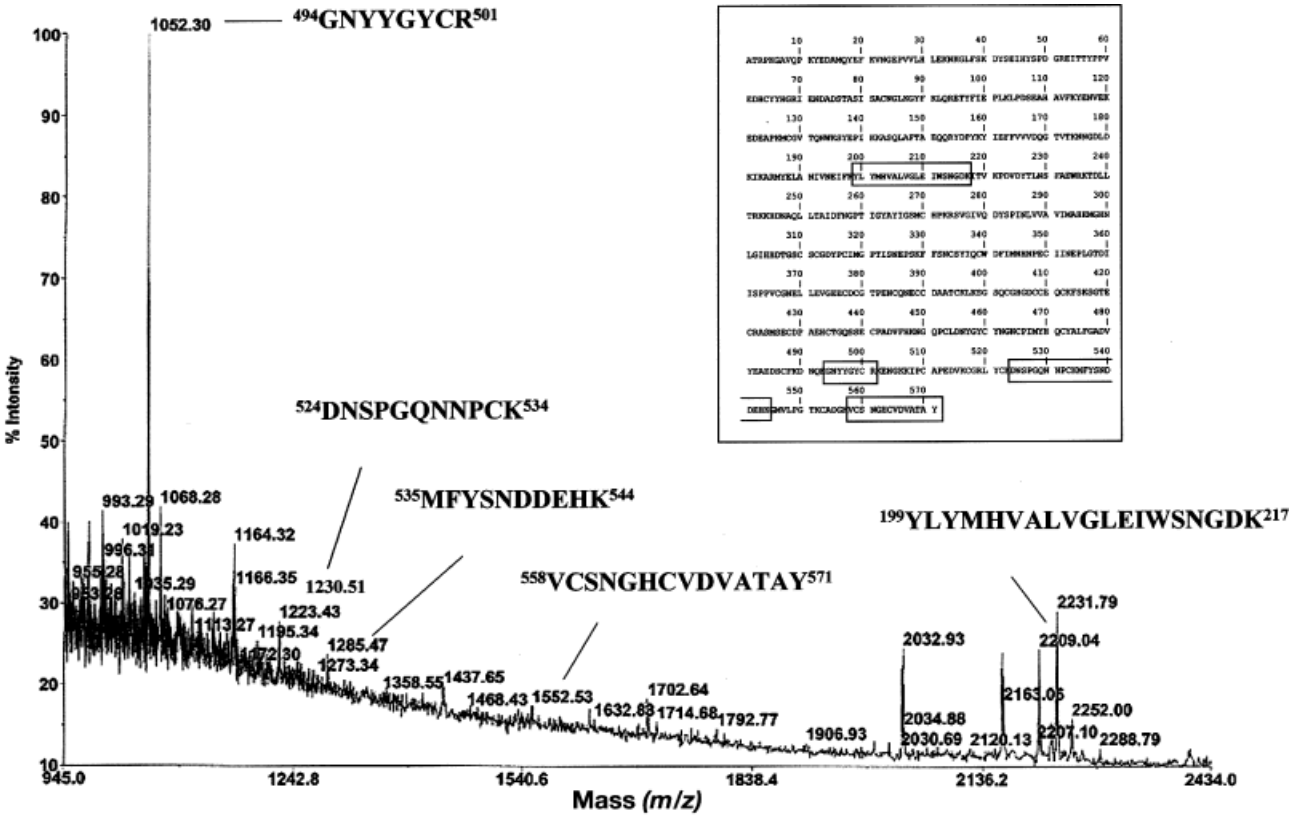


Figure 4. Unprocessed MALDI-TOF peptide mass fingerprint of a tryptic digest of the CBB-stained SDS polyacrylamide gel-separated protein band of RP HPLC fraction 15, showing the proposed peptide sequence assignments made by the Protein Prospector program (v.3.4.1) and their location within the sequence of the metalloproteinase jararhagin precursor from *B. jararaca* venom (SWISS-PROT accession code P30431).

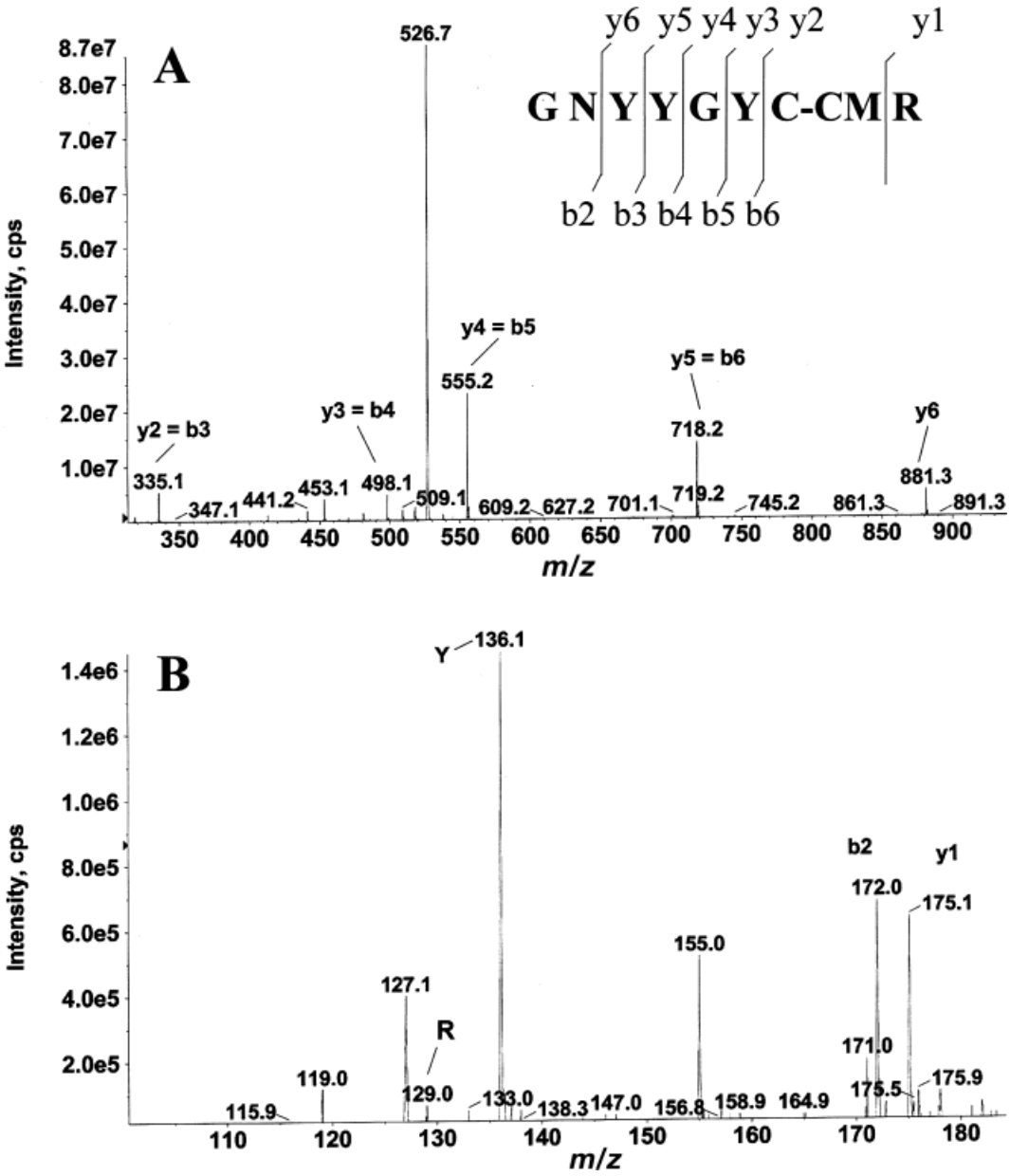
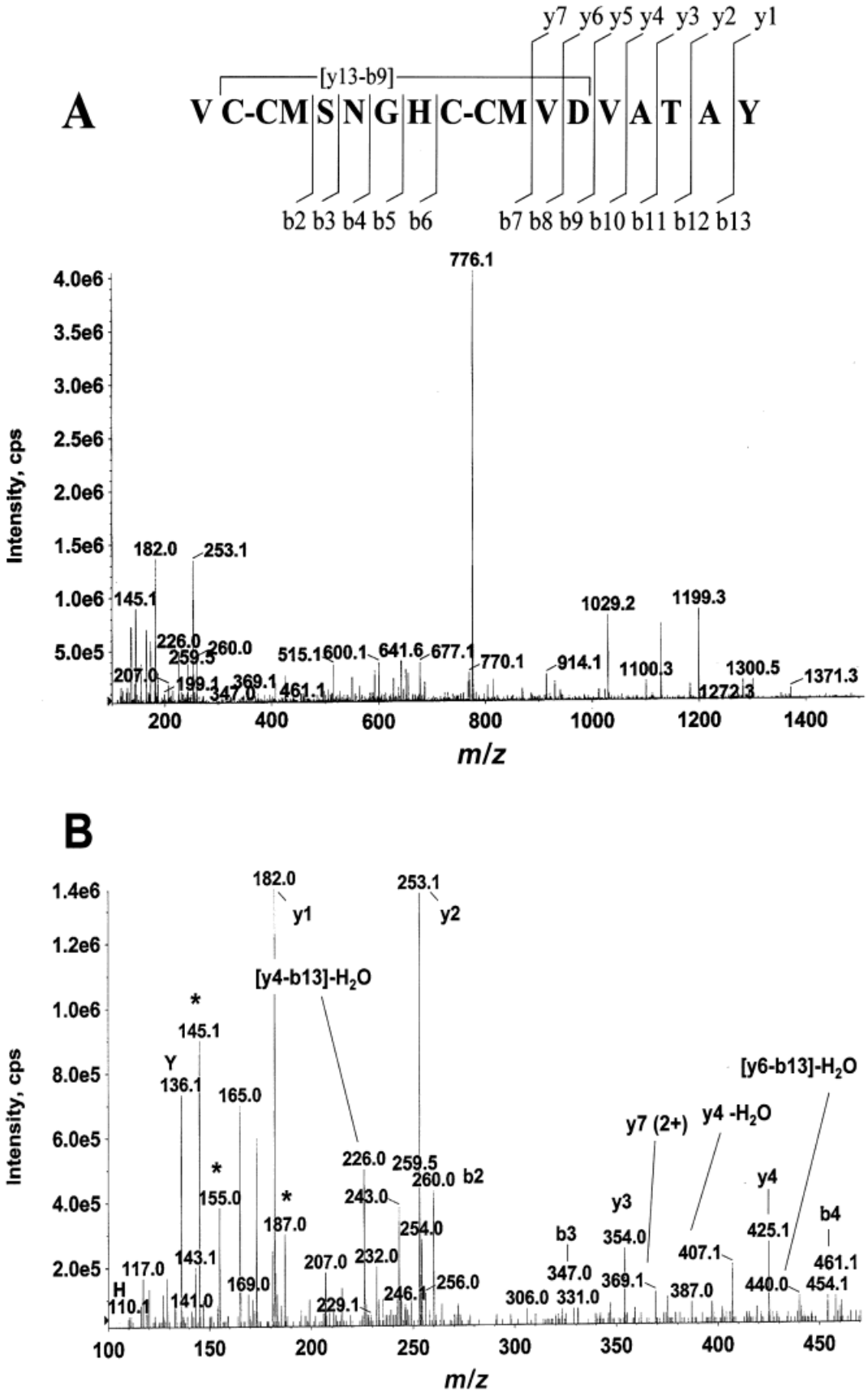


Figure 5. High (A) and low (B) mass range of the MS/MS spectra of the doubly-charged monoisotopic ion at *m/z* 526.7 (ion *m/z* 1052.3 in Fig. 4) showing the b and y ions from which the corresponding amino acid sequence (displayed in panel A) was deduced. The immonium ions of arginine at *m/z* 129.0 and tyrosine at *m/z* 136.1 are labelled "R" and "Y", respectively, in panel B. CM-C, carbamidomethyl cysteine.

mass. Using the operating conditions specified in Section 2.6, singly-charged y1–y6 ions and singly- and doubly-charged b ions encompassing b2–b13 were generated (Fig. 6 B-D) This information, along with some a-ions and internal fragment ions of the a, b and b-H_2O series, unambiguously confirmed the identity of *m/z* 776.1 (2+) as VCSNGHCVDVATAY, as suggested by the mass fingerprint approach.

In addition to PIII hemorrhagin, the venom of *S. barbouri* also contains other metalloproteases, which eluted in RP HPLC fractions 13 and 14 (Fig. 2, Table 1). They are characterized by the presence of a free cysteine and 3 disulfide bonds. Their *N*-terminal sequences and molecular masses suggest that they may be isoenzymes. The *N*-terminal sequences of these *S. barbouri* metalloproteases are identical to that of atrolysin E (EC 3.4.24.44) from the



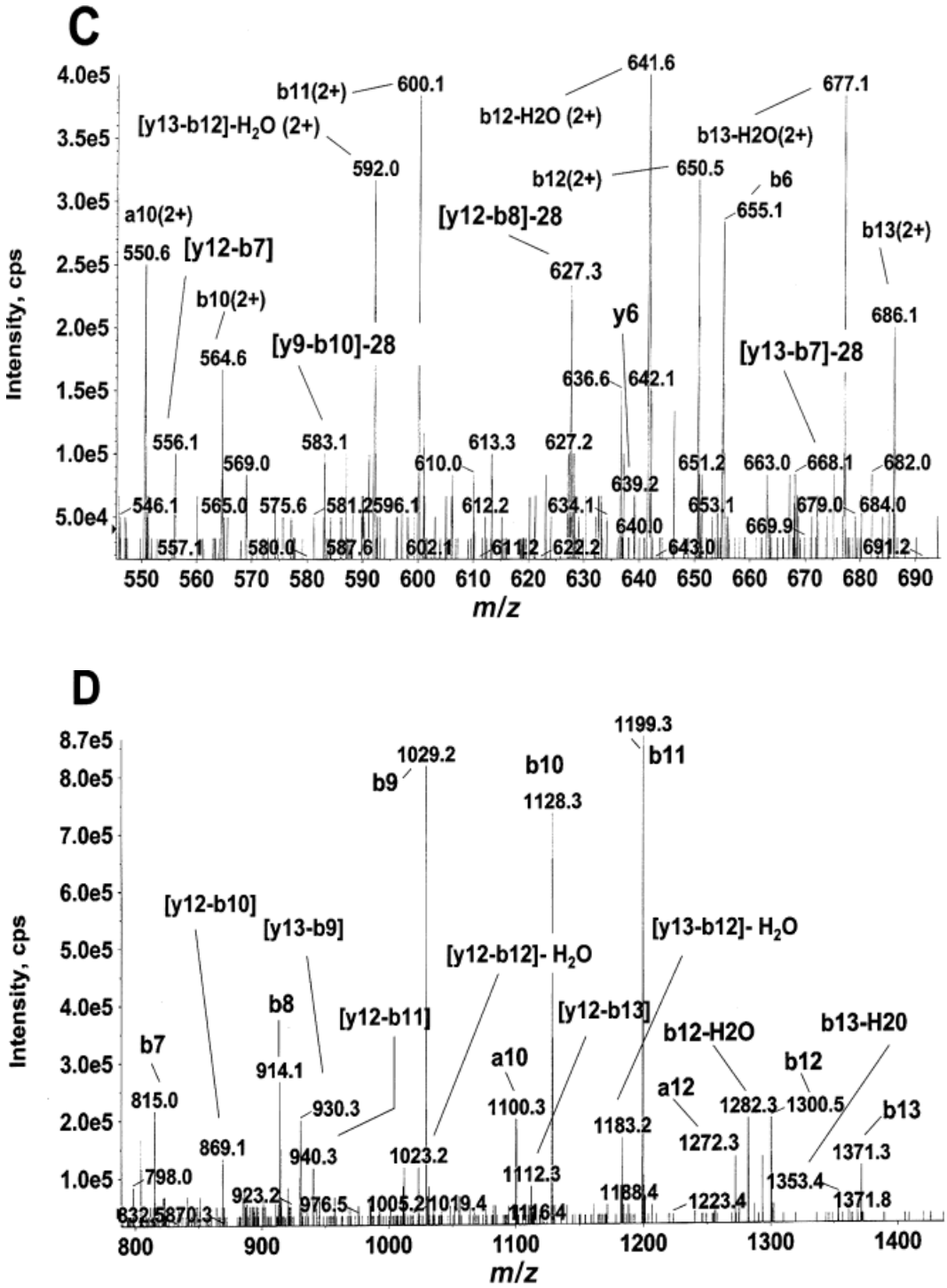


Figure 6. (A) Full MS/MS spectrum and sequence assignment of the doubly-charged monoisotopic ions at *m/z* 776.1 (ion *m/z* 1552.5 in Fig. 4). (B–D) Details of different mass ranges and selected daughter ions used to confirm the sequence VCSNGHCVDVATAY are labelled. The nomenclature for sequence-specific ("b" and "y" ions) proposed by Biemann [40] is used. Internal ions are defined by their y- and b-type cleavages at the two peptide bonds specified in between square brackets. The immonium ions of His and Tyr at *m/z* 110.1 and 136.1 are labelled H and Y, respectively in panel B. Asterisks in panel B mark the singly protonated internal amino acyl ion [y4-b12]-28 AT or [y3-b13]-28 TA (*m/z* 145.1), the internal ion [y4-b12]-H_2O AT or [y3-b13]-H_2O TA (*m/z* 155.0), and the internal ion [y7-b9]-28 VD or [y6-b9]-28 DV at *m/z* 187.1. CM-C, carbamidomethyl cysteine.

western diamondback rattlesnake *C. atrox* (P34182). At present, it is not clear whether the metalloproteases in fractions 13 and 14 belong to the PI, PII or PIII class. PI metalloproteinases (20–30 kDa) are single-domain proteins with relatively weak hemorrhagic activity. The class PII metalloproteinases (30–60 kDa) contain a disintegrin domain at the *c*-terminus of a metalloproteinase domain structurally similar to that present in class PI metalloproteinases. However, the fact that fractions 4 and 5 contained disintegrins, antagonists of integrin receptors which are released in venoms by proteolytic processing of PII metalloproteinases [8, 32], and the lack of disintegrin/cysteine-rich fragments, which are proteolytically derived from PIII metalloproteinases [33, 34], strongly support the PI/PII origin of the proteins in fractions 13 and 14.

3.3 A protein of the CRISP family

The protein isolated in HPLC fraction 8 displays strong similarity to other venom cysteine-rich secretory proteins (CRISPs) from a number of *Viperidae* and *Elapidae* snakes [35–37] in terms of its *N*-terminal sequence, molecular mass, and the number of cysteine residues. The occurrence of CRISPs in the venoms of snakes from different continents has been realized only very recently [37]. Our results support the hypothesis that CRISPs may represent a widely distributed protein family among snake venoms. Three venom CRISPs, piscivorin (*A. piscivorus piscivorus, Viperidae*, USA), ophanin (*Ophiophagus hannah, Elapidae*, Southeast Asia), and catrin-2 (*C. atrox, Viperidae*, USA and Mexico) showed mild but significant inhibition of rat tail arterial smooth muscle evoked by high $K+$ concentration [37]. Other CRISPs such as ablomin (*Agkistrodon blomhoffi*, *Viperidae*, Japan), triflin (*Trimeresurus flavoviridis, Viperidae*, Japan), and latisemin (*Laticauda semifasciata, Elapidae*, Southeast Asia), inhibited depolarization-induced contraction of rat tail arterial smooth muscle, showing properties of L-type Ca^{2+} channel blocking toxins [35]. CRISPs pseudechetoxin (*Pseudechis australis, Elapidae*, North and Central Australia) and pseudecin (*Pseudechis porphyriacus, Elapidae*, South Australia), blocked olfactory and retinal cyclic nucleotide-gated ion channel currents [36]. Snake venom CRISPs have highest similarity to helothermine, a toxin from the salivary secretion of the Mexican beaded lizard (*Heloderma horridum horridum*) [38]. Helothermine targets a number of ion channels, including voltage-gated Ca^{2+} and K^+ channels, and ryanodine receptors, and its physiological effects in rodents are lethargy, rear limb, paralysis, hypothermia, and death [39]. The biological effects of the *S. barbouri* snake venom CRISP requires detailed investigations.

4 Concluding remarks

The accelerated evolution of snake venom gland proteins has been documented. This phenomenon, which may have adaptative consequences, represents a serious handicap for the identification of proteins by mass peptide fingerprinting, even if the proteins belong to close evolutionary related species. Nevertheless, most snake venom protein families are characterized by their cysteine content. We have applied a cysteine mapping approach, in conjunction with *N*-terminal sequencing and mass spectrometric data to classify the proteins of *S. barbouri* venom into defined protein families. Although the lack of snake genome sequences is an additional drawback for the identification of venom proteins by MALDI-TOF mass fingerprinting, MS/MS fragmentation of selected ions yielded sufficient amino acid sequence information, derived from an almost complete series of b- and/or y-ions and internal ions, to unambiguously identify an homologue protein from a *S. barbouri* related snake venom protein. Our results show that the venom proteome of the pigmy rattlesnake *S. barbouri* is composed of proteins belonging to a few known protein families, and support the hypothesis that CRISP molecules may represent a widely distributed protein family among certain snake venoms.

This work was financed by grant BMC2001-3337 from the Dirección General de Enseñanza Superior e Investigación Científica, Madrid, Spain (to JJC). PJ and LS are recipients of a pre-doctoral fellowship (FPI, formación de personal investigador) from the Spanish Ministerio de Ciencia y Tecnología, and an I3P contract, respectively.

Received April 30, 2003
Revised July 10, 2003
Accepted September 1, 2003

5 References

[1] Markland, F. S., *Toxicon* 1998, *36*, 1749–1800.
[2] Fry, B. G., *Toxicon* 1999, *38*, 11–32.
[3] Ménez, A. (Ed.), *Perspectives in Molecular Toxinology*, John Wiley and Sons, Chichester, UK 2002.
[4] Harvey, A. L., *Toxicon* 2001, *39*, 15–26.
[5] Mizuno, H., Fujimoto, Z., Koizumi, M., Kano, H. *et al.*, *Nat. Struct. Biol.* 1997, *4*, 438–441.
[6] Mizuno, H., Fujimoto, Z., Koizumi, M., Kano, H. *et al.*, *J. Mol. Biol.* 1999, *289*, 103–112.
[7] Fukuda, K., Mizuno, H., Atoda, H., Morita, T., *Biochemistry* 2000, *39*, 1915–1923.
[8] McLane, M. A., Marcinkiewicz, C., Vijay-Kumar, S., Wierzbicka-Patynowski, I. *et al.*, *Proc. Soc. Exp. Biol. Med.* 1998, *219*, 109–119.
[9] Adler, M., Lazarus, R. A., Dennis, M. S., Wagner, G., *Science* 1991, *253*, 445–448.

[10] Saudek, V., Atkinson, R. A., Pelton, J. T., *Biochemistry* 1991, *30*, 7369–7372.

[11] Senn, H., Klaus, W., *J. Mol. Biol.* 1993, *232*, 907–925.

[12] Smith, K. J., Jaseja, M., Lu, X., Williams, J. A. *et al.*, *Int. J. Pept. Protein Res.* 1996, *48*, 220–228.

[13] Scarborough, R. M., Rose, J. W., Hsu, M. A., Phillips, D. R. *et al.*, *J. Biol. Chem.* 1991, *266*, 9359–9362.

[14] Marcinkiewicz, C., Calvete, J. J., Marcinkiewicz, M. M., Raida, M. *et al.*, *J. Biol. Chem.* 1999, *274*, 12468–12473.

[15] Calvete, J. J., Moreno-Murciano, M. P., Theakston, R. D. G., Kisiel, D. G., Marcinkiewicz, C., *Biochem. J.* 2003, *372*, 725–734.

[16] Marcinkiewicz, C., Calvete, J. J., Vijay-Kumar, S., Marcinkiewicz, M. M. *et al.*, *Biochemistry* 1999, *38*, 13302–13309.

[17] Calvete, J. J., Fox, J. W., Agelan, A., Niewiarowski, S. *et al.*, *Biochemistry* 2002, *41*, 2014–2021.

[18] Tani, A., Ogawa, T., Nose, T., Nikandrov, N. N. *et al.*, *Toxicon* 2002, *40*, 803–813.

[19] Moura-Da-Silva, A. M., Theakston, R. D. G., Crampton, J. M., *J. Mol. Evol.* 1996, *43*, 263–269.

[20] Deshimaru, M., Ogawa, T., Nakashima, K., Nobuhisa, I. *et al.*, *FEBS Lett.* 1996, *397*, 83–88.

[21] Harvey, A. L., Bradley, K. N., Cochran, S. A., Rowan, E. G. *et al.*, *Toxicon* 1998, *36*, 1635–1640.

[22] Dunn, R. D., Broady, K. W., Biochim. Biophys. Acta 2001, *1533*, 29–37.

[23] Neves-Ferreira, A. G. C., Perales, J., Fox, J. W., Shannon, J. D. *et al.*, *J. Biol. Chem.* 2002, *277*, 13129–13137.

[24] Thompson, J. D., Higgins, D. J., Gibson, T. J., *Nucleic Acid Res.* 1994, *22*, 4673–4680.

[25] Gil, F., Ramón-Maiques, S., Marina, A., Fita, I. *et al.*, *Acta Crystallogr.* 1999, *D55*, 1350–1352.

[26] Uriarte, M., Marina, A., Ramón-Maiques, S., Fita, I. *et al.*, *J. Biol. Chem.* 1999, *274*, 16295–16303.

[27] Mann, K., Farias, C. M. S. A., Gallego del Sol, F., Santos, C. F. *et al.*, *Eur. J. Biochem.* 2001, *268*, 4414–4422.

[28] Friedman, M., Krull, L. H., Cavins, J. F., *J. Biol. Chem.* 1970, *245*, 3868–3871.

[29] Henschen, A., in: Wittmann-Liebold, B., Salnikow, J., Erdman, V. A. (Eds.), *Advanced Methods in Protein Microsequence Analysis*, Berlin, Germany 1986, pp. 244–255.

[30] Hager, J. W., *Rapid Commun. Mass Spectrom.* 2002, *16*, 512–526.

[31] Hati, R., Mitra, P., Sarker, S., Bhattacharyya, K. K., *Crit. Rev. Toxicol.* 1999, *29*, 1–19.

[32] Kini, R. M., Evans, H. J., *Toxicon* 1992, *30*, 265–293.

[33] Shimokawa, K.-I., Shannon, J. D., Jia, L.-G., Fox, J. W., *Arch. Biochem. Biophys.* 1997, *343*, 35–43.

[34] Calvete, J. J., Moreno-Murciano, M. P., Sanz, L., Jürgens, M. *et al.*, *Protein Sci.* 2000, *9*, 1365–1373.

[35] Yamazaki, Y., Koike, H., Sugiyama, Y., Motoyoshi, K. *et al.*, *Eur. J. Biochem.* 2002, *269*, 2708–2715.

[36] Yamazaki, Y., Brown, R. L., Morita, T., *Biochemistry* 2002, *41*, 11331–11337.

[37] Yamazaki, Y., Hyodo, F., Morita, T., *Arch. Biochem. Biophys.* 2003, *412*, 133–141.

[38] Morrissette, J., Krätzschmar, J., Haendler, B., El-Hayek, R. *et al.*, *Biophys. J.* 1995, *68*, 2280–2288.

[39] Mochca-Morales, J., Martín, J. B. M., Possani, L. D., *Toxicon* 1990, *28*, 299–309.

[40] Biemann, K., *Annu. Rev. Biochem.* 1992, *61*, 977–1010.

3.2. ARTÍCULO 2: Molecular cloning of disintegrin-like transcript BA-A5 from a *Bitis arietans* venom gland cDNA library: a putative intermediate in the evolution of the long-chain disintegrin Bitistatin

J Mol Evol (2006) xx:1–11
DOI: 10.1007/s00239-005-0268-z

JOURNAL OF MOLECULAR EVOLUTION

Molecular Cloning of Disintegrin-like Transcript BA-5A from a *Bitis arietans* Venom Gland cDNA Library: A Putative Intermediate in the Evolution of the Long-Chain Disintegrin Bitistatin

Paula Juárez,[1] Simon C. Wagstaff,[2] Jenny Oliver,[2] Libia Sanz,[1] Robert A. Harrison,[2] Juan J. Calvete[1]

[1] Instituto de Biomedicina de Valencia, CSIC, Jaime Roig 11, 46010 Valencia, Spain
[2] Alistair Reid Venom Research Unit, Liverpool School of Tropical Medicine, Pembroke Place, Liverpool L3 5QA, UK

Received: 10 November 2005 / Accepted: 1 March 2006 [*Reviewing Editor:* Dr. Bryan Grieg Fry]

Abstract. We report the cloning and sequence analysis of BA-5A from a venom gland cDNA library of the puff adder, *Bitis arietans*, that encodes a novel ECD-disintegrin-like domain. BA-5A is a unique PII disintegrin. It contains the 16 cysteine residues that are conserved in all known disintegrin-like domains of ADAM proteins and snake venom metalloproteinases but lacks the cysteine-rich domain. These features suggest that BA-5A may represent an intermediate in the evolutionary pathway of the long disintegrin bitistatin and that removal of the cysteine-rich domain and loss of the PIII-specific disulfide bond were separate events along the structural diversification pathway of disintegrins, the former predating the latter. The protein family composition of the *Bitis arietans* venom, as determined by combination of reversed-phase HPLC and proteomic analysis, was as follows: Zn^{2+}-metalloproteinase (38.5%), serine proteinase (19.5%), disintegrin (17.8%), C-type lectin-like (13.2%), PLA_2 (4.3%), Kunitz-type inhibitor (4.1%), cystatin (1.7%), and unknown (0.9%). BA-5A could not be detected in the venom proteome of *Bitis arietans*. The occurrence of this very low-abundance (< 0.05%) or nonexpressed disintegrin transcript indicates a hitherto unrecognized structural diversity of this protein family. Whether BA-5A plays a physiological role or represents an orphan protein which could eventually evolve a role in the adaptation of snakes to changing ecological niches and prey habits deserves further investigation.

Correspondence to: Juan J. Calvete; *email:* jcalvete@ibv.csic.es

Key words: Snake venomics — Venom proteome — *Bitis arietans* — cDNA cloning — Disintegrin evolution

Introduction

Snake venom proteins play a number of adaptative roles: immobilizing, paralyzing, killing, digesting prey, and deterring competitors. Venoms of vipers and rattlesnakes (subfamilies Viperinae and Crotalinae of Viperidae) contain protein toxins that initiate hemorrhage and other toxins that prevent the function of normal hemostatic responses to arrest bleeding, including platelet aggregation and the effectors resulting from initiation of the coagulation cascade. These proteins can be grouped into a few major protein families, including enzymes (serine proteinases, Zn^{2+}-metalloproteases, L-amino acid oxidase, group II phospholipases A_2 [PLA_2]) and proteins without enzymatic activity (disintegrins, C-type lectins, natriuretic peptides, cysteine-rich secretory proteins [CRISP] toxins, nerve growth factors, cystatin, and Kunitz-type protease inhibitors) (Markland 1998; Juárez et al. 2004; Fry and Wüster 2004; Fry 2005). Current evidence suggests that many of these toxin gene families were recruited from proteins with normal, nontoxic, physiological function into the venom proteome early in the evolution of advanced snakes at the base of the colubroid tree (Fry and Wüster 2004; Fry 2005; Fry et al. 2005). On the other hand, the

existence in the same venom of a diversity of isoforms of proteins of the same family differing from each other in their pharmacological effects results from paralogous genes originated by gene duplications and accelerated Darwinian evolution (Menez 2002; Tani et al. 2002; Moura da Silva et al. 1996). A fast coevolutionary arms race between snakes and their prey as a driving mechanism in the evolution of venom proteins has been hypothesized (Daltry et al. 1996).

The diversity of biological activities of proteins sharing the same general structural scaffold has been investigated in a number of venom protein groups, including the disintegrins (Menez 2002; Tani et al. 2002; Gomis-Rüth 2003; Tsai et al. 2004; Lu et al. 2005a, b; Calvete et al. 2005; Calvete 2005). Disintegrins are small (40–100 amino acids), cysteine-rich polypeptides that selectively block the function of integrin receptors (Calvete et al. 2005; Calvete 2005). Currently, the disintegrin family can be conveniently divided into five groups according to their length (40–100 residues) and the number (four to eight) of disulfide bonds (Calvete et al. 2003). The first group includes short disintegrins, composed of 41–51 residues and 4 disulfide bonds. The second group is formed by the medium-sized disintegrins, which contain about 70 amino acids and 6 disulfide bonds. The third group includes long disintegrins, with an ~84-residue polypeptide cross-linked by 7 disulfide bonds. The fourth subfamily of disintegrins groups the disintegrin-like domains derived from PIII snake venom metalloproteinases (SVMPs). PIII disintegrins are modular proteins containing an N-terminal disintegrin-like domain of about 100 amino acids including 16 cysteine residues involved in the formation of 8 disulfide bonds and a C-terminal 110- to 120-residue cysteine-rich domain cross-linked by 6 disulfides (Calvete et al. 2000a). The disintegrin-like domains of PIII SVMPs molecules lack the integrin-binding motif present in the integrin-binding loops of many PII disintegrins (RGD, KGD, WGD, VGD, MLD, etc.), which is replaced by XXECD sequences. Unlike the PII (short, medium, and long) and PIII disintegrins, which are single-chain molecules, the fifth group is composed of homo- and heterodimers. Dimeric disintegrins contain subunits of about 67 residues, with 10 cysteines involved in the formation of 4 intrachain disulfide bonds and 2 interchain cystine linkages (Calvete et al. 2000b; Bilgrami et al. 2004, 2005). Bilitoxin-1 represents another homodimeric disintegrin comprising disulfide-bonded polypeptides, each containing 15 cysteinyl residues (Nikai et al. 2000).

Disintegrins proper (small, medium-sized, long, and dimeric) are released in the venoms of various vipers by proteolytic processing of larger mosaic PII metalloprotease precursors (Kini and Evans 1992) or synthesized from short-coding mRNAs (Okuda et al. 2002; Sanz et al. 2006). The antagonistic activity of PII disintegrins toward integrin receptors ($\alpha_1\beta_1$, $\alpha_3\beta_1$, $\alpha_4\beta_1$, $\alpha_4\beta_7$, $\alpha_5\beta_1$, $\alpha_6\beta_1$, $\alpha_6\beta_4$, $\alpha_7\beta_1$, $\alpha_8\beta_1$, $\alpha_9\beta_1$, $\alpha_v\beta_3$, $\alpha_v\beta_1$, $\alpha_v\beta_3$, $\alpha_{IIb}\beta_3$) depends on the appropriate pairing of cysteine residues, which determine the conformation of the mobile inhibitory loop which protrudes 14–17 Å from the protein core and harbors the active tripeptide at its apex (Moreno-Murciano et al. 2003; Monleón et al. 2005; and references cited therein). In contrast, the PIII disintegrin-like domains contain an extra disulfide bond between CysXIII (within the XXECD motif) and CysXVI, which may restrain the conformation of their ECD-loop. Disintegrin-like/cysteine-rich domains containing RSECD or MSECD sequences have been reported to inhibit integrin $\alpha_2\beta_1$-mediated collagen-induced platelet aggregation (reviewed by Calvete et al. 2005).

Functional diversification between disintegrins is mainly due to amino acid substitutions within the active loop, whereas structural diversification was driven through a disulfide bond engineering mechanism involving the selective loss of pairs of cysteine residues engaged in the formation of disulfide bonds (Calvete et al. 2003). The great sequence and structural diversity exhibited by the different subfamilies of disintegrins strongly suggests that disintegrins, like toxins from other venoms (Duda and Palumbi 1999; Kordis et al. 2002; Ohno et al. 2002), have evolved rapidly by adaptative evolution. The accelerated evolution of toxins may be linked to adaptation to the environment, including feeding habits (Okuda et al. 2001).

Research on disintegrins not only is relevant for understanding the biology of viper venom toxins, but also provides information on new structural determinants involved in integrin recognition that may be useful in basic and clinic research. To understand the genomic basis of the accelerated evolution of disintegrins, and the molecular mechanism underlying their structural diversification, we have searched for messages encoding disintegrins in a *Bitis arietans* venom gland library. We report the cloning of a cDNA encoding a novel ECD disintegrin-like domain containing the 16-cysteine scaffold conserved in all known ADAM domains but lacking the cysteine-rich domain. Proteomic analysis failed to detect this unique disintegrin-like domain in the venom proteome of *Bitis arietans*, which is hypothesized to represent an intermediate in the evolutionary pathway of the long disintegrin bitistatin.

Materials and Methods

cDNA Library Synthesis

Total RNA was extracted from pooled venom glands of two specimens of *B. arietans*. The vipers were sacrificed 3 days after venom extraction, when toxin gene transcription rates are at a peak (Paine et al. 1992). Glands were homogenized under liquid N_2 and total RNA (Trizol; Invitrogen) and then mRNA (PolyATract; Promega)

was extracted following the manufacturers' guidelines. The cDNA library was constructed according to the manufacturer's protocol from the mRNA using the SMART cDNA library construction kit (Clontech, USA), which involved an initial reverse transcriptase step followed by a PCR step of 27 cycles which yielded cDNA varying from 250 to 2500 bp. The latter was size fractionated (500–2500 bp), inserted into the λTriplEx2 vector, and packaged into λ phage using Gigapack III Gold Packaging Extract (Stratagene). The resultant amplified cDNA library contained 8×10^9 plaque-forming units/ml. This material was boiled for 5 min prior to being used as targets of polymerase chain reaction (PCR) amplification.

cDNA Cloning and Sequencing

A forward primer, 5′-CCAAATCCAGC/TCTCCAAAATG-3′, and a reverse primer, 5′-TTCCAG/TCTCCATTGTTGG/TTTA, complementary to highly conserved 5′- and 3′-noncoding regions of cDNA encoding for elegantin-2a from *Trimeresurus elegans* (GenBank accession number AB059572), elegantin-1a from *T. elegans* (GenBank accession number AB059571), and HR2a from *Trimeresurus flavoviridis* (accession code AY037808) were synthesized. The PCR protocol, using venom gland cDNA as template, included an initial denaturation step at 95°C for 6 min followed by 35 cycles of denaturation (1 min at 94°C), annealing (1 min at 55°C), and extension (1 min at 74°C) and a final extension for 7 min at 72°C. AmpliTaq Gold (Roche), a highly processive 5′–3′ DNA polymerase that lacks 3′–5′ exonuclease activity, was used. The inclusion of water-only controls with each PCR reaction allowed us to monitor and prevent crossover contamination. The amplicons were subcloned into the TA cloning vector, pCR 2.1-TOPO (Invitrogen, Gröningen, The Netherlands), and used to transform chemically competent *E. coli* cells (TOP10F'; Invitrogen) under ampicillin selection, resulting in numerous (> 50) colonies. Plasmid DNA was extracted (Mini-spin prep kit; Qiagen, Hilden, Germany) from four randomly selected colonies and digested with *Bam*H1 and *Xho*1 at 37°C to select plasmids containing inserts of the predicted size for DNA sequencing.

Isolation and Characterization of Venom Proteins

Venom was collected by snake biting on a parafilm-wrapped jar and pooled from 15 wild-caught *B. arietans* specimens (Ghana) of different ages and of both sexes and maintained in the herpetarium of the Liverpool School of Tropical Medicine. Venom was lyophilized and stored at 4°C in a dark bottle until used. For reverse-phase HPLC separation, 2.2 mg of the crude venom was dissolved in 100 µl of 5% acetonitrile and 0.1% trifluoroacetic acid (TFA). Insoluble material was removed by centrifugation in an Eppendorf centrifuge (Hamburg, Germany) at 13,000 *g* for 10 min at room temperature. Soluble proteins were separated with an ETTAN LC HPLC system (Amersham Biosciences) using a Lichrospher RP100 C18 column (250 × 4 mm, 5-mm particle size; Merck, Darmstadt, Germany) eluted at 1 ml/min with a linear gradient of 0.1% TFA in water (solution A) and in acetonitrile (solution B), first isocratically (5% B) for 5 min, followed by linear gradients of 5–45% B for 120 min and 45–70% B for 20 min. Protein detection was at 215 nm, and peaks were collected manually. The isolated protein fractions were analyzed by SDS-PAGE, N-terminal sequencing (using an Applied Biosystems Procise 492 sequencer), and matrix-assisted laser-desorption ionization-time-of-flight mass spectrometry (MALDI-TOF-MS).

MALDI-TOF Mass Spectrometry

For mass determination and quantitation of sulfhydryl groups and disulfide bonds, the purified proteins (1 µg in 2 µl of 100 mM ammonium bicarbonate, pH 8.3, containing 5 M guanidinium hydrochloride) were incubated either with 10 mM iodoacetamide for 1 h at room temperature or with 10 mM DTT for 15 min at 65°C, followed by the addition of a fivefold molar excess of iodoacetamide over-reducing agent and incubation for 1 h at room temperature. The reaction mixtures were freed from reagents using a C18 Zip-Tip pipette (Millipore) after activation with 70% acetonitrile and equilibration in 0.1% TFA. Following protein adsorption and washing with 0.1% TFA, the proteins were eluted onto the MALDI-TOF plate with 1 µl of 70% acetonitrile and 0.1% TFA and subjected to mass spectrometric analysis. The molecular masses of the native and the reduced and carbamidomethylated lectins were determined by MALDI-TOF mass spectrometry using an Applied Biosystems Voyager DE-PRO instrument operating at 25 kV accelerating voltage in the linear mode, and using 3,5–dimethoxy-4-hydroxycinnamic acid (sinapinic acid) saturated in 70% acetonitrile and 0.1% TFA as the matrix. The mass calibration standard consisted of a mixture of the following proteins, whose isotope-averaged molecular mass, as daltons, are given in parentheses: bovine insulin (5734.5), *E. coli* thioredoxin (11,674.5), and horse apomyoglobin (16,952.6).

The number of free cysteine residues (N_{SH}) was determined using eq. (1):

$$N_{SH} = (M_{IA} - M_{NAT})/57.05 \quad (1)$$

where M_{IA} is the mass of the denatured but nonreduced protein incubated in the presence of iodoacetamide, M_{NAT} is the mass of the native protein, and 57.05 is the mass increment due to the carbamidomethylation of one thiol group.

The number of total cysteine residues (N_{Cys}) was derived using eq. (2):

$$N_{Cys} = [(M_{CM} - M_{IA})/58.05] + N_{SH} \quad (2)$$

where M_{CM} is the mass of the reduced and carbamidomethylated protein, and 58.05 is the mass increment due to the carbamidomethylation of a cysteine residue, which prior to reduction was involved in the formation of a disulfide bond.

Finally, the number of disulfide bonds ($N_{S\text{-}S}$) was calculated from eq. (3):

$$N_{S-S} = (N_{cys} - N_{SH})/2 \quad (3)$$

All mass values in eqs. (1)–(3) are in daltons.

Sequence Similarity Searches and Phylogenetic Analysis

Amino acid sequence similarity searches were carried out against a nonredundant protein databank using the program PSI-BLAST (Altschul et al. 1997) accessible at http://www.ncbi.nlm.nih.gov/BLAST. Program MEGA (Molecular Evolutionary Genetic Analysis; http://www.megasoftware.net) (Kumar et al. 2001) was employed for inferring phylogenies (evolutionary trees) from a multiple alignment of disintegrin sequences.

In-Gel Enzymatic Digestion and Mass Fingerprinting

Reverse-phase HPLC-separated fractions containing blocked N-termini or heterogeneous N-terminal sequences were analyzed by SDS-PAGE. All the SDS-PAGE-separated, Coomassie brilliant blue-stained protein bands were excised from the gels and subjected to automated digestion with sequencing-grade bovine pancreas trypsin (Roche) at a final concentration of 20 ng/µl in 50 mM ammonium bicarbonate, pH 8.3, using a ProGest digestor (Genomic Solutions) following the manufacturer's instructions. Digestions were conducted after reduction with DTT (10 mM for 15 min at 65°C) and carbamidomethylation with iodoacetamide (50 mM for 60 min at room temperature). The tryptic peptide mixtures were dried in a SpeedVac, and the samples were dissolved in 5 µl of 50%

acetonitrile and 0.1% TFA, then subjected to mass fingerprinting. When necessary, the digestion mixtures were diluted with 0.1% TFA to a final acetonitrile concentration of < 10% and were freed from reagents using a C18 Zip-Tip pipette tip (Millipore) as above.

For mass fingerprinting analysis, 0.85 μl of the digests was spotted onto a MALDI-TOF sample holder, mixed with an equal volume of a saturated solution of α-cyano-4-hydroxycinnamic acid (Sigma) in 70% acetonitrile containing 0.1% TFA, dried, and analyzed with an Applied Biosystems Voyager-DE Pro MALDI-TOF mass spectrometer, operated in delayed extraction and reflector modes. Database searches were constrained to a mass tolerance of 100 ppm. A tryptic peptide mixture of *Cratylia floribunda* seed lectin (SwissProt accession code P81517) prepared and previously characterized in our laboratory was used as the mass calibration standard (mass range, 450–3300 Da).

Collision-Induced Dissociation by Tandem Mass Spectrometry

For peptide sequencing, the protein digest mixture was subjected to electrospray ionization tandem mass spectrometric (MS/MS) analysis using a QTrap mass spectrometer (Applied Biosystems) equipped with a nanospray source (Protana, Denmark). Doubly charged ions selected after Enhanced Resolution MS analysis were fragmented using the Enhanced Product Ion with Q_0 trapping option at 250 amu/s across the entire mass range. For MS/MS experiments, Q1 was operated at unit resolution, the Q1-to-Q2 collision energy was set at 35 eV, the Q3 entry barrier was 8 V, the LIT (linear ion trap) Q3 fill time was 250 ms, and the scan rate in Q3 was 1000 amu/s. Collision-induced dissociation (CID) spectra were interpreted manually or using the on-line form of the MASCOT program at http://www.matrixscience.com.

Database Accesion Codes

The cDNA sequence clone BA-5A has been deposited with the EMBL Nucleotide Sequence Data Bank (http://www.ebi.ac.uk/) under accession code AM117393.

Results and Discussion

Disintegrins have been reported to date in the venoms of a number of genera from the subfamilies Crotalinae (*Agkistrodon*, *Bothrops*, *Calloselasma*, *Crotalus*, *Deinagkistrodon*, *Gloydius*, *Lachesis*, *Protobothrops*, *Sistrurus*, and *Trimeresurus*) and Viperinae (*Bitis*, *Cerastes*, *Daboia*, *Echis*, *Eristocophis*, *Macrovipera*, and *Vipera*), which represent about 50% of the classified genera of family Viperidae (http://www.embl-heidelberg.de/~uetz/families/Viperidae.html). For the majority of the venoms of snakes from the other half of the genera of Viperidae, toxin compositional analyses have not been addressed. Hence, disintegrins may represent a widely distributed venom protein family. Most of the venoms of snakes from the examined genera contain medium-sized, dimeric, and/or short disintegrins. Long disintegrins have been reported so far only in three species, *Gloydius halys brevicaudus* (salmosin-3) (Park et al. 1998), *Agkistrodon bilineatus* (bilitoxin-1) (Nikai et al. 2000), and *Bitis arietans* (bitistatin) (Shebuski et al. 1989). Based on biochemical and phylogenetic analyses, we have proposed a model for the structural diversification of disintegrins, in which the long disintegrins derive from the disintegrin-like/cysteine-rich domains of a PIII metalloprotease by deletion of the cysteine-rich domain and the PIII-specific disulfide bond between cysteine XIII and cysteine XVI (Calvete et al. 2003). To check this hypothesis we have analyzed a *Bitis arietans* venom gland cDNA library looking for novel disintegrin-coding transcripts.

BA-5A, a Putative Intermediate in the Evolution of Bitistatin

The full-length clone B5-A5 was amplified from *B. arietans* venom gland polyadenylated RNA using primers for the highly conserved 5′ and 3′ noncoding region of known disintegrins. The deduced amino acid sequence of the full-length open reading frame, assembled from four identical overlapping cDNA sequences from a single PCR-amplified band of about 1600 bp, is shown in Fig. 1 and reveals the presence of a signal peptide, a pro-domain, a metalloproteinase domain, and a disintegrin-like domain. A similar multidomain structure has been reported in a number of PII and PIII snake venom metalloproteinases (SVMPs), the precursors of disintegrin proper and disintegrin-like/cysteine-rich (DC) fragments, respectively (Lu et al. 2005b; Fox and Serrano 2005). However, B5-A5 departs from the canonical structures of PII- and PIII-SVMP precursors in two main features: (i) it contains a disintegrin-like (ECD) instead of a disintegrin proper (RGD) domain, and (ii) it lacks a cysteine-rich domain (Fig. 2). Thus, full-length BA-5A must be regarded as a PII metalloprotease with a PIII-like disintegrin domain.

Disintegrin-like domains are thought to represent ancestral molecules of the long and medium disintegrins (Calvete et al. 2003) and are distinguished from the disintegrins proper by the length of their polypeptides (100 vs. 80 residues, respectively), which include16 cysteine residues, and by the expression of an XCD motif in lieu of the typical RGD (KGD, VGD, WGD, MGD, MLD, etc.) integrin-binding motif of the disintegrins proper (Calvete et al. 2005; Calvete 2005). The disulfide bonding pattern of 14 of the 16 cysteines of disintegrin-like domains is conserved in the structure of the long disintegrins (Calvete et al. 1997, 2000) (Fig. 3). The other two cysteines are located within the XCD motif and in the C-terminal region, and are engaged in the formation of a disintegrin-like domain-specific disulfide bond. In the structure of BA-5A the two extra cysteines are residues 69, within a ^{66}RSECD^{70} motif, and Cys95 (Fig. 2).

Snake venom PIII disintegrins evolved from the extracellular domains of cell membrane-anchored ADAM (a disintegrin and metalloprotease) molecules

```
|----- Signal sequence ----->                                          |-- Pro-
 M   M   Q   V   L   L   V   T   I   C   L   A   V   F   P   Y   Q   G      18
ATG ATG CAA GTT CTC TTA GTA ACT ATA TGC TTA GCA GTT TTT CCA TAT CAA GGG      54

domain ------>
 S   S   I   I   L   E   S   G   N   V   N   D   Y   E   V   V   Y   P      36
AGC TCT ATA ATC CTG GAA TCT GGG AAC GTT AAT GAT TAT GAA GTA GTG TAT CCA     108

 Q   K   V   T   A   L   P   K   G   A   V   Q   Q   A   E   Q   K   Y      54
CAA AAG GTC ACT GCA CTG CCC AAA GGA GCA GTT CAG CAG GCT GAG CAA AAG TAT     162

 E   D   A   M   Q   Y   E   F   E   V   N   G   Q   P   V   V   L   H      72
GAA GAT GCC ATG CAA TAT GAA TTT GAA GTG AAT GGA CAG CCA GTG GTC CTT CAC     216

 L   E   K   N   K   D   L   F   S   E   D   Y   S   E   T   H   Y   S      90
CTA GAA AAA AAT AAA GAT CTT TTT TCA GAA GAT TAC AGT GAG ACT CAT TAT TCA     270

 P   D   G   K   E   I   T   T   N   P   P   I   E   D   H   C   Y   Y     108
CCT GAT GGC AAA GAA ATT ACA ACA AAC CCT CCA ATT GAG GAT CAC TGC TAT TAT     324

 H   G   R   I   Q   N   D   A   H   S   T   A   S   I   S   A   C   N     126
CAT GGA CGG ATC CAG AAT GAT GCT CAC TCA ACT GCA AGC ATC AGT GCA TGC AAT     378

 G   L   K   G   H   F   K   L   R   G   E   T   Y   L   I   E   P   L     144
GGT TTG AAA GGA CAT TTC AAG CTT CGA GGG GAG ACG TAT TTA ATT GAA CCC TTG     432

 K   I   P   D   S   E   A   H   A   V   Y   K   Y   E   N   I   E   K     162
AAG ATT CCT GAC AGT GAA GCC CAT GCA GTC TAC AAA TAT GAG AAC ATA GAA AAA     486

 E   D   D   A   P   K   M   C   G   V   T   Q   T   N   W   E   S   D     180
GAG GAT GAT GCC CCC AAA ATG TGT GGG GTA ACC CAG ACT AAT TGG GAA TCA GAT     540

                                                  |-- Metalloproteinase -->
 E   P   I   K   E   A   S   Q   L   V   A   T   S   D   Q   Q   R   Y     198
GAG CCC ATC AAA GAG GCC TCT CAG TTA GTT GCT ACG TCT GAT CAA CAA AGA TAC     594

 Y   D   H   F   R   Y   I   K   Y   F   I   V   V   D   H   R   M   V     216
TAT GAC CAC TTC AGA TAC ATT AAG TAT TTC ATA GTT GTG GAC CAC AGA ATG GTT     648

 E   K   Y   N   G   N   L   R   T   I   R   R   R   I   Y   Q   L   V     234
GAG AAA TAC AAT GGT AAT TTA AGA ACG ATA AGA AGA AGA ATA TAT CAA CTT GTC     702

 N   I   L   N   E   I   Y   L   P   W   N   I   R   A   P   L   V   G     252
AAC ATT TTA AAT GAG ATA TAC TTA CCT TGG AAT ATT CGT GCA CCA CTG GTT GGC     756

 I   E   F   W   N   Q   R   D   L   I   N   V   T   S   S   A   P   Y     270
ATA GAA TTT TGG AAC CAA AGA GAT TTG ATT AAT GTG ACG TCA TCA GCA CCA TAT     810

 T   L   D   L   F   G   K   W   R   A   S   D   L   L   N   R   K   I     288
ACT TTG GAC TTA TTT GGA AAA TGG AGA GCA TCA GAT TTG CTG AAT CGC AAA ATA     864

 H   D   Y   T   H   L   L   T   A   I   V   F   V   E   Q   I   L   G     306
CAT GAT TAT ACT CAC TTA CTC ACG GCC ATT GTT TTT GTT GAA CAA ATA TTA GGA     918

 M   A   H   I   A   T   M   C   H   S   E   L   S   V   G   L   V   Q     324
ATG GCT CAC ATA GCC ACC ATG TGC CAT TCA GAA CTT TCT GTA GGA CTT GTT CAG     972

 D   Y   M   P   S   E   H   V   V   A   A   I   M   V   H   E   M   G     342
GAT TAT ATG CCA TCA GAG CAC GTG GTT GCA GCT ATA ATG GTC CAC GAG ATG GGT    1026

 H   N   L   G   I   S   H   D   E   K   Y   C   N   C   G   A   D   S     360
CAT AAC CTG GGC ATT AGT CAT GAT GAA AAA TAC TGT AAT TGT GGT GCT GAC TCA    1080

 C   I   M   Y   P   Q   I   S   I   P   P   P   V   Y   F   S   N   C     378
TGC ATT ATG TAT CCT CAG ATA AGC ATT CCA CCT CCT GTG TAT TTC AGC AAT TGT    1134

 S   W   E   Q   Y   Q   N   F   L   T   I   Y   K   P   D   C   T   L     396
AGT TGG GAG CAA TAT CAG AAT TTT CTT ACT ATT TAT AAA CCA GAT TGC ACT CTC    1188

                                  |-------- Disintegrin-like domain ---->
 I   R   P   S   R   T   D   I   V   S   P   P   V   C   G   N   D   I     414
ATC AGA CCC TCG AGA ACT GAT ATT GTT TCA CCT CCA GTT TGT GGA AAT GAT ATT    1242

 L   E   Q   G   E   E   C   D   C   G   S   P   E   K   C   Q   D   P     432
TTG GAG CAG GGA GAA GAA TGC GAC TGT GGC TCT CCT GAA AAG TGT CAA GAT CCG    1296

 C   C   D   A   A   S   C   K   L   H   S   W   I   E   C   E   F   G     450
TGC TGC GAT GCT GCA TCA TGT AAA CTA CAC TCA TGG ATA GAG TGT GAA TTT GGA    1350

 E   C   C   D   Q   C   R   F   K   P   A   G   T   E   C   R   G   I     468
GAG TGT TGC GAC CAA TGC AGA TTT AAG CCA GCA GGA ACA GAA TGC CGG GGA ATA    1404

 R   S   E   C   D   L   P   E   Y   C   T   G   Q   S   V   D   C   P     486
AGA AGT GAG TGT GAC CTG CCT GAA TAC TGC ACT GGC CAA TCT GTT GAC TGT CCT    1458

 I   D   H   F   H   R   N   G   K   P   C   L   N   N   N   G   A   E     504
ATA GAT CAC TTC CAT AGG AAT GGA AAA CCA TGC CTA AAC AAC AAT GGA GCG GAA    1512

 K   G   E   F   Q   H   T   G   G   R   Y   *                           516
AAG GGC GAA TTC CAG CAC ACT GGC GGC CGT TAC TAG                          1548
```

Fig. 1. DNA and deduced amino acid sequences of the *Bitis arietans* BA-5A clone. The nucleotide sequences are numbered in the 5′–3′ direction from the initial codon ATG to the stop codon TAG. The signal sequence and the predicted mature protein sequences are underlined and in boldface, respectively. The N-termini of the signal peptide, pro-domain, metalloproteinase, and disintegrin-like domain are labeled. The positions of the Cys-switch site (KMCGV), the Zn^{+2}-binding motif (HEMGHNLGISH) within the metalloprotease domain, and the RSECD sequence in the disintegrin-like domain are shaded.

Fig. 2. Amino acid sequence alignment of the *Bitis arietans* BA-5A clone, the long disintegrins bitistatin, its D1, D2, D3, and DC1 isoforms, and salmosin-3, and the disintegrin-like domains of selected PIII metalloproteases. Swiss-Prot/TrEMBL (http://us.expasy.org/sprot/) accession codes are in parentheses. The sequences of clones Eo-00006 and Eo-10c1 are from Juárez et al. (2006). Cysteine residues are shaded. Residues of BA-5A different from bitistatin are marked with asterisks on top of the bitistatin sequence, and those of the bitistatin isoforms departing from bitistatin P17497 are in italics and underlined. The Cys30/Tyr mutation in bitistatin DC1 is labeled with a romb. The RGD integrin-binding motif of bitistatin and salmosin-3 and the topological equivalent tripeptides in BA-5A and the PIII disintegrin-like domains are highlighted in boldface. The highly conserved N-terminal part of the cysteine-rich domains of the latter are underlined and labeled "Cys-rich."

after mammals and reptiles diverged (Moura da Silva et al. 1996). Phylogenetic analysis, in conjunction with biochemical and genetic data, support the model depicted in Fig. 3B, by which the structural diversification of the disintegrin family occurred through the successive loss of disulfide bonds (Calvete et al. 2003). In the phylogenetic tree of disintegrins, BA-5A segregates into the clade formed by the PIII and the long disintegrins (Fig. 3A). From a structural point of view, BA-5A represents an intermediate species between a PIII disintegrin-like molecule and a long disintegrin.

In all PIII-SVMP precursor open reading frames characterized to date the disintegrin-like (D) domain is followed by a C-terminal cysteine-rich (C) domain (Lu et al. 2005b; Fox and Serrano 2005). We have hypothesized that the concerted loss of the disintegrin-like-specific, integrin loop-constraining CysXIII-CysXVI linkage and the cysteine-rich domain may have paved the way for the emergence of the single-domain PII disintegrins (Calvete et al. 2003, 2005) (Fig. 3). The finding of BA-5A, a PIII disintegrin-like domain lacking the C-terminal cysteine-rich domain, calls for a revision of the proposed scheme for the evolution of long disintegrins from PIII SVMPs. Thus, the occurrence of BA-5A supports the view that removal of the cysteine-rich domain and loss of the PIII-specific disulfide bond are separate events, the former predating the latter (Fig. 3B). Hence, the updated model depicted in Figs. 3B and 4 includes BA-5A as an intermediate in the evolutionary pathway leading to the emergence of the long disintegrin bitistatin through gene duplication and the stepwise removal from the duplicated PIII metalloproteinase

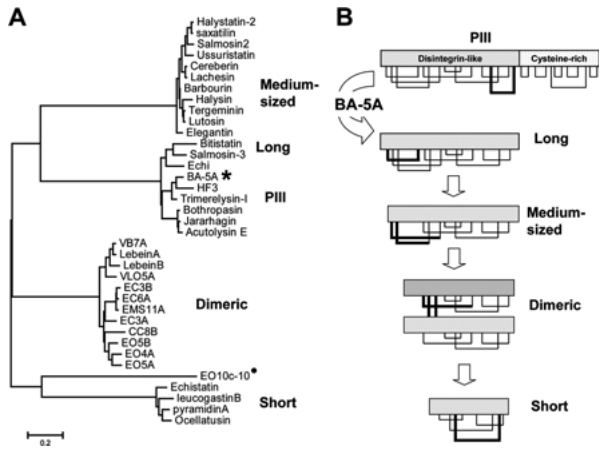


Fig. 3. **A** Cladogram for the multiple sequence analysis of selected representative from the different snake venom disintegrin subfamilies. The non-venom-secreted disintegrin-like BA-5A described in this work is labeled with an asterisk. The non-venom-secreted ocellastusin precursor, EO10c-10, described in the accompanying paper (Juárez et al. 2006), is labeled with a filled circle. The tree represents the minimum evolutionary distance estimated through neighbor joining using maximum likelihood distances. Maximum parsimony produced a similar topology. The length of the horizontal scale bar represents 20% divergence. For primary references on the analyzed disintegrins, consult Calvete et al. (2003). **B** Scheme of the domain organization, disulfide bond patterns, and proposed evolutionary pathway from the PIII disintegrin-like/cysteine-rich proteins to short disintegrins. Structural features (the cysteine-rich domain of PIII disintegrin-like molecules, and class-specific disulfides) lost along the disintegrin diversification pathway are highlighted with thick lines. In the proposed model for the evolutionary divergence of disintegrins, BA-5A is hypothesized to represent an intermediate structure between a PIII disintegrin-like protein and the long disintegrin bitistatin (Fig. 4).

precursor gene of the cysteine-rich domain and the PIII disintegrin-like domain-specific cystine linkage, followed by the emergence of the RGD motif.

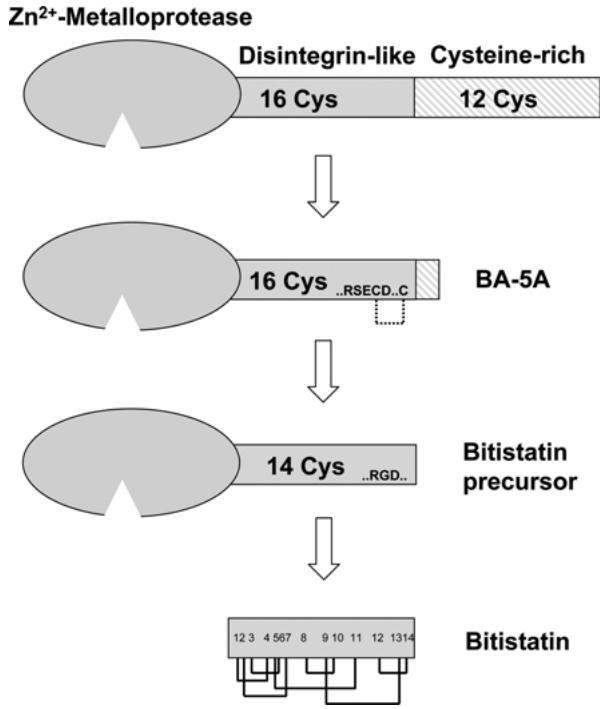


Fig. 4. Illustration of the proposed evolutionary pathway of the long disintegrin bitistatin, which includes the stepwise removal of the cysteine-rich domain from a PIII metalloproteinase precursor gene, yielding the BA-5A gene, loss of the PIII disintegrin-like domain-specific CysXIII-CysXVI linkage, and emergence of the RGD motif. The mature bitistatin structure, showing its seven disulfide bonds as connecting lines, is depicted.

BA-5A Was Not Detected in the B. arietans *Venom Proteome*

PIII-SVMPs can undergo proteolysis/autolysis during secretion or in the venom to produce a biologically active, two-domain product comprising of a disintegrin-like (D) and a C-terminal cysteine-rich (C) domain and termed a DC-fragment. The consensus from several functional studies on a number of PIII-SVMPs and DC-fragments suggests that the disintegrin-like and cysteine-rich domains are likely to play a role in the composite activity of the modular PIII-SVMPs, by targeting the toxin to the cell surface $\alpha_2\beta_1$ integrin and, as a processed product, by blocking the binding of collagen ligands to the integrin (Lu et al. 2005b; Calvete et al. 2005; Fox and Serrano 2005). Most of the $\alpha_2\beta_1$ integrin-blocking DC-fragments exhibit RSECD (i.e., atrolysin A from *Crotalus atrox*) (Jia et al. 1997) like B5-A5 or MSECD (catrocollastatin-C from *Crotalus atrox* (Zhou et al. 1995; Shimokawa et al. 1997), jararhagin-C from *Bothrops jararaca* (Moura da Silva et al. 2001; Zigrino et al. 2002), and alternagin-C from *Bothrops alternatus* (Souza et al. 2000)) motifs. These data suggested an $\alpha_2\beta_1$ integrin inhibitory activity for B5-A5 (^{66}RSECD^{70} motif; Fig. 2). To test this hypothesis, we sought to isolate and study this unique disintegrin from the pooled venom of 15 wild-caught specimens of Ghana *B. arietans* snakes. To this end, venom proteins were separated by reverse-phase HPLC (Fig. 5) and the isolated protein peaks were submitted to N-terminal sequencing and mass spectrometric analysis (Table 1). However, no evidence for the presence in the venom of the isolated B5-A5 disintegrin-like domain was found. The expected molecular mass of the isolated B5-A5 disintegrin-like domain should range between the isotope-averaged molecular mass of full-length BA-5A (^{1}VSPPV....GGRY^{111}), which is 12180 Da, assuming that all cysteines are involved in disulfide bonding, and that of a fully oxidized processed molecule encompassing residues ^{6}CGNDI...GKPC^{93} (minimal mass), which is 9727 Da. No HPLC peak fulfilled these criteria. Peaks 1–6 showed N-terminal sequences that depart in just two residues from that of BA-5A displayed in Fig. 2 and may hence represent close relatives of BA-5A. However, the mass difference of 813 Da between the reduced and carbamidomethylated and the native BA-1-6 molecules (Table 1) clearly indicated that each of these disintegrins contained (813/58) = 14 cysteine residues and not 16, as would be expected for BA-5A.

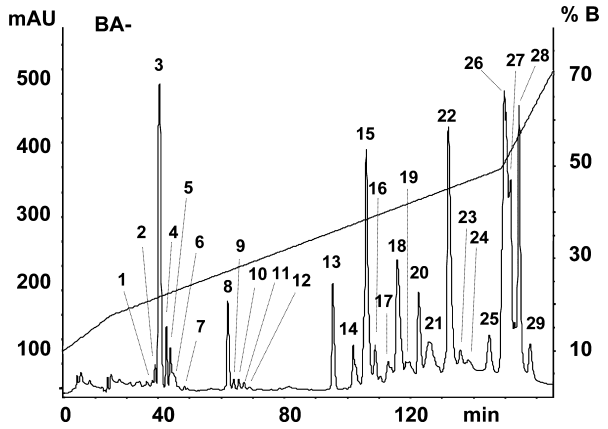


Fig. 5. Reverse-phase HPLC separation of the *Bitis arietans* venom proteome. Protein fractions whose identities were characterized by combination of N-terminal sequencing and mass spectrometry (Table 1) are labeled.

HPLC peaks 26–28 corresponded to SVMPs exhibiting the same N-terminal sequence and sharing internal sequences as determined by MS/MS analysis (Table 1). Their molecular masses clearly indicated that Ba-26/27 (23 kDa) belong to the PI class of metalloproteases, whereas Ba-28 (46 kDa) represents a PIII-SVMP. Further structural data are needed to determine whether Ba-28 and Ba-26/27 are actually the full-length and the processed metalloproteinase domain of the same gene product. Peak Ba-29 also corresponded to a metalloprotease of high molecular mass, which shares with 26–28 an internal peptide sequence. The current limited sequence data (Fig. 5, Table 1) indicate that all these SVMPs are clearly distinct proteins from BA-5A. The fact that BA-5A was undetectable in the *B. arietans* venom proteome

Table 1. Assignment of the reversed-phase isolated fractions of *Bitis arietans* venom to protein families by N-terminal Edman sequencing, MALDI-TOF mass spectrometry, and collision-induced fragmentation by nESI-MS/MS of selected peptide ions from in-gel digested protein bands

HPLC fraction	N-Terminal sequencing	Isotope-averaged MALDI-TOF mass (±0.2%)	Peptide ion m/z	z	MS/MS-derived sequence	Protein family
1	SPPVCGNKILEQGED	8,824.4[b] 9,637[a]				Disintegrin
2	SPPVCGNKILEQGED	9,008.4[b] 9,821[a]				Disintegrin (Bitistatin 1-83) (P17497)
3	SPPVCGNKILEQGED	8,991.4[b] 9,804[a]				Disintegrin Bitistatin D1 (AAY43681)
4	VSPPVCGNKILEQGED	9,018.9[b] 9,831[a]				Disintegrin
5	SPPVCGNKILEQGED	8,851.1[b] 9,664[a]				Disintegrin
6	SPPVCGNKILEQGED	9,291.0[b] 10,104[a]				Disintegrin
7	N.D.	9,571.4[b] 10,383[a]				Disintegrin
8	Blocked	6,942 (7,398[a])	533.9	2	FKTPEECR	Kunitz-type inhibitor
			396.2	2	TPEECR	
9,10	N.D.	7 kDa[c]	533.9	2	FKTPEECR	Kunitz-type inhibitor
			396.2	2	TPEECR	
11,12	N.D.	17,622	461.8	2	XFCEXNK	Unknown
			530.3	2	XPC(233.1)PXVK	
13	SLVEFGQMIQEETEK	13,903				PLA_2
14	IPGGLSPRDVTDPDV	13,219				Cystatin (P08935)
15	VIGGDECNINEHRSL M	42 kDa[c]	467.1	2	YFCXSSR	Ser-proteinase
			530.6	2	XFDYSVCR	
			608.1	2	VFDYTDWXR	
			757.1	2	VIGGDECNINEHR	
		m 30 kDa[c]	696.5	3	FHCAGTXXNKEWVXTAAR	Ser-proteinase
			530.6	2	XFDYSVCR	
		m 17 kDa[c]	510.8	2	NPFICKSR	C-Type lectin-like
16	VIGGDECNINEHRSL	[28,667, 30,182]				Ser-proteinase
17	VIGGDECNINEHRSL	29,911	608.1	2	VFDYTDWXR	Ser-proteinase
			757.1	2	VIGGDECNINEHR	
18	VIGGAECNINEHRSL	30608				Ser-proteinase
19	VIGGAECNINEHRSL	30 kDa[c]				Ser-proteinase
20	DPGCLPDWSSYKVFK GCLPDWSSYKGHCYK	29,959				C-Type lectin-like
21	N.D.	30 kDa	620.8	2	YEKSWAEAEK	C-type lectin-like CTL-5 (AAQ01208)
			850.2	2	FVYDAWIGLRDESK	
			510.8	2	NPFICKSR	
			699.2	2	CFGLDVHTEYR	
22	DPGCLPDWSSYKVFKKV DEGCLPDWSSYKGHCYK	29,716				C-Type lectin-like (bitiscetin) Q7LZK5
23,24	N.D.	30 kDa	510.8	2	NPFICKSR	C-Type lectin-like
25	DFQCPSEWSAYGQHCY DQDCLSGWSFYFETCY	33,382				C-Type lectin-like CTL-5 (AAQ01208)
26,27	(sid)PIKYINVIVVAD-QRLVTYYKGELNKIT	23,310	753.9	2	SASDTXHSFVTWR	PI metalloproteinase
			855.9	2	(171.2) EMWSNGDXCTVTK	
			472.3	2	FISTHNPK	
28	(sid) PIKYINVIVVAD-QRLVTYYKGELNKIT	46,585	753.9	2	SASDTXHSFVTWR	PIII metallo-proteinase
			855.9	2	(171.2) EMWSNGDXCTVTK	
			472.3	2	FISTHNPK	
29	N.D.	>90 kDa	753.9	2	SASDTXHSFVTWR	Metalloproteinase

Note. X, Ile or Leu. Unless stated otherwise, for MS/MS analyses, cysteine residues were carbamidomethylated. N.D., not determined.
[a]MALDI-TOF mass of the reduced and carboamidomethylated molecule.
[b]Molecular mass determined by electrospray-ionization mass spectrometry.
[c]Apparent molecular mass determined by SDS-PAGE after sample reduction with β-mercaptoethanol.

supports the view that it may represent a nontranslated transcript or a very low-abundance protein (<0.05% of the total venom proteins). Alternatively, BA-5A may be translated but not secreted into the venom, it may be secreted into the venom of some but not all individuals, or the protein may exhibit a temporal expression pattern over the lifetime of the snake. Clearly, further work is needed to clarify this point.

Table 2. Overview of the relative occurrence of proteins (as a percentage of the total HPLC-separated proteins) of the different families in the venoms of Ghana *Bitis arietans*

Protein family	%
Disintegrin	17.8
Kunitz-type inhibitor	4.1
PLA_2	4.3
Cystatin	1.7
Serine proteinase	19.5
C-Type lectin-like	13.2
Zn^{2+}-metalloproteinase	38.5
Unknown	0.9

The B. arietans *Venom Proteome*

The snake venom proteome of *B. arietans* was investigated using a combination of protein chemical (reverse-phase HPLC and N-terminal sequencing) and proteomic (MALDI-TOF mass determination and fingerprinting of in-gel digested protein bands, and sequence analysis of peptide ions by MS/MS) analyses. Hence, HPLC fractions showing single N-terminal sequence and molecular mass could be straightforwardly assigned to a given protein family by BLAST database search (Table 1). Those HPLC fractions with either blocked or uninterpretable (heterogeneous) N-termini were subjected to SDS-PAGE and the resulting protein bands were characterized by MALDI-TOF mass fingerprinting followed by MS/MS of selected ions. In general the product ion spectra did not match with any known protein upon automatic database search with MASCOT. However, in most cases a sufficiently long string of sequence could be manually deduced, which was then used for sequence similarity analysis using BLAST. This strategy allowed us to characterize protein fractions which represent less than 0.05% of the total venom proteins. An overview of the relative contribution of proteins from different families to the total venom proteome of *B. arietans* is shown in Table 2. The data are in line with the view that although snake venom contains complex mixtures of pharmacologically active molecules, the toxins can be grouped into a few major protein families (Juárez et al. 2004; Bazaa et al. 2005). Thus, the *B. arietans* venom proteome is quite simple, with its major proteins ($>10\%$) belonging to only four major and three minor protein families (Table 2). Noteworthy, disintegrins represent over 17% of the total venom proteins. This figure is 0.1–8% in the Tunisian snakes *Macrovipera lebetina* (6%), *Cerastes vipera* ($<1\%$), and *Cerastes cerastes* (8%) (Bazaa et al. 2005) and in the North American snakes of the genus *Sistrurus*, *Sistrurus miliarius barbouri* (8%), *S. catenatus catenatus* (2.5%), *S. catenatus tergeminus* (2–4%), and *S. catenatus edwardsii* (0.1–0.9%) (L. Sanz, S. P. MacKessy, H. L. Gibbs, & J. J. Calvete, unpublished results). On the other hand, no evidence for the presence of natriuretic peptides, nerve growth factors, myotoxins, CRISPs, and L-amino acid oxidase, which form part of the toxin arsenal of other advanced snakes of the Viperidae family (Juárez et al. 2004; Fry and Wüster 2005; Fry 2005; Fry et al. 2005; Bazaa et al. 2005), was found in the *B. arietans* venom. Knowledge of the relative amounts of different toxin families in a given venom might be relevant for generating immunization protocols to elicit toxin-specific antibodies showing greater specificity and effectiveness than the conventional methods relying on the immunization of large mammals with whole venom.

The molecular masses determined for HPLC peaks BA-2 and BA-3 fit accurately with the calculated isotope-averaged molecular masses of the long disintegrin bitistatin (SwissProt/TrEMBL accession code P17497) and its isoform D1 (AAY43681), respectively. On the other hand, we did not find evidence, by either N-terminal sequencing or mass spectrometry, for the presence of the bitistatin isoforms D2, D3, and DC1 whose nucleotide sequences deduced from venom gland cDNAs of *B. arietans* were deposited on February 2005 in the EMBL/GenBank/DDBJ databases by J. Oliver, R. G. D. Theakston, and R. A. Harrison, under accession codes AAY43682, AAY43683, and AAY43684, respectively (Fig. 2). These results indicate that BA-1 to -6 correspond to novel isoforms of bitistatin or bitistatin-like disintegrins, and support the view of the existence of greater disintegrin sequence variability in the genome/transcriptome than in the venom proteome of *B. arietans*. It is also notable that the bitistatin isoform DC1 includes a Cys30/Tyr mutation (Fig. 2), which indicates that this gene product, if expressed, would have a nonnative disulfide arrangement. Interestingly, three other DNA messengers coding for disintegrin domains exhibiting the same Cys/Tyr mutation have been reported, including MD2 from *Deinagkistrodon acutus* (Tsai et al. 2000) and clones Eo-00006 and Eo-10c1 from *Echis ocellatu*s (Juárez et al. 2006) (Fig. 2). The latter *E. ocellatus* messengers are not expressed in the venom and appear to represent key intermediates in the evolutionary pathway leading to the emergence of the short disintegrin ocellatusin from a short-coding dimeric disintegrin precursor (Juárez et al. 2006). All this evidence points to the existence of very low-abundance, nonexpressed, or non-venom-secreted disintegrin transcripts, which may play a hitherto unrecognized physiological function, or may simply represent orphan molecules which eventually could serve as scaffolds for evolving novel biological activities, perhaps becoming functional proteins of relevance for the adaptation of snakes to changing ecological niches and prey habits.

Acknowledgments. This work was financed by Grant BFU2004-01432 from the Ministerio de Educación y Ciencia, Madrid, Spain (to J.J.C.). P.J. and L.S. are recipients of a predoctoral fellowship (FPI; formación de personal investigador) from the Spanish Ministerio de Educación y Ciencia and a postdoctoral I3P contract, respectively. R.A.H. and S.C.W. were funded by the Wellcome Trust, and J.O. by the Medical Research Council.

References

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Bazaa A, Marrakchi N, El Ayeb M, Sanz L, Calvete JJ (2005) Snake venomics: comparative analysis of the venom proteomes of the Tunisian snakes *Cerastes cerastes*, *Cerastes vipera* and *Macrovipera lebetina*. Proteomics 5:4223–4235

Bilgrami S, Tomar S, Yadav S, Kaur P, Kumar J, Jabeen T, Sharma S, Sinhg TP (2004) Crystal structure of schistatin, a disintegrin homodimer from saw-scaled viper (*Echis carinatus*) at 2.5 Å resolution. J Mol Biol 341:829–837

Bilgrami S, Yadav S, Sharma S, Perbandt M, Betzel C, Singh TP (2005) Crystal structure of the disintegrin heterodimer from saw-scaled viper (*Echis carinatus*) at 1.9 Å resolution. Biochemistry 44:11058–11066

Calvete JJ (2005) Structure-function correlations of snake venom disintegrins. Curr Pharm Des 11:829–835

Calvete JJ, Schrader M, Raida M, McLane MA, Romero A, Niewiarowski S (1997) The disulfide bond pattern of bitistatin, a disintegrin isolated from the venom of the viper *Bitis arietans*. FEBS Lett 416:197–202

Calvete JJ, Moreno-Murciano MP, Sanz L, Jürgens M, Schrader M, Raida M, Benjamin DC, Fox JW (2000a) The disulfide bond pattern of catrocollastatin C, a disintegrin-like/cysteine-rich protein isolated from *Crotalus atrox* venom. Protein Sci 9:1365–1373

Calvete JJ, Jürgens M, Marcinkiewicz C, Romero A, Schrader M, Niewiarowski S (2000b) Disulfide bond pattern and molecular modelling of the dimeric disintegrin EMF–10, a potent and selective integrin α5β1 antagonist from *Eristocophis macmahoni* venom. Biochem J 345:573–581

Calvete JJ, Moreno-Murciano MP, Theakston RDG, Kisiel DG, Marcinkiewicz C (2003) Snake venom disintegrins: novel dimeric disintegrins and structural diversification by disulfide bond engineering. Biochem J 372:725–734

Calvete JJ, Marcinkiewicz C, Monleón D, Esteve V, Celda B, Juárez P, Sanz L (2005) Snake venom disintegrins: evolution of structure and function. Toxicon 45:1063–1074

Daltry JC, Wüster W, Thorpe RS (1996) Diet and snake venom evolution. Nature 379:537–540

Duda TF Jr, Palumbi SR (1999) Developmental shifts and species selection in gastropods. Proc Natl Acad Sci USA 96:6820–6823

Fox JW, Serrano SMT (2005) Structural considerations of the snake venom metalloproteinases, key members of the M12 reprolysin family of metalloproteinases. Toxicon 45:969–985

Fry BG (2005) From genome to "venome": molecular origin and evolution of the snake venom proteome inferred from phylogenetic analysis of toxin sequences and related body proteins. Genome Res 15:403–420

Fry BG, Wüster W (2004) Assembling an arsenal: origin and evolution of the snake venom proteome inferred from phylogenetic analysis of toxin sequences. Mol Biol Evol 21:870–883

Fry BG, Vidal N, Norman JA, Vonk FJ, Scheib H, Ramjan SFR, Kuruppu S, Fung K, Hedges SB, Richardson MK, Hodgson WC, Ignjatovic V, Summerhayes R, Kochva E (2005) Early evolution of the venom system in lizards and snakes. Nature 439:584–588

Gomis-Rüth FX (2003) Structural aspects of the Metzincin clan of metalloendopeptidases. Mol Biotechnol 24:157–202

Harrison RA, Oliver J, Hasson SS, Bharati K, Theakston RDG (2003) Novel sequences encoding venom C-type lectins are conserved in phylogenetically and geographically distinct *Echis* and *Bitis* viper species. Gene 315:95–102

Jia L-G, Wang X-M, Shannon JD, Bjarnason JB, Fox JW (1997) Function of disintegrin-like/cysteine-rich domains of atrolysin A. Inhibition of platelet aggregation by recombinant protein and peptide antagonists. J Biol Chem 272:13094–13102

Juárez P, Sanz L, Calvete JJ (2004) Snake venomics: characterization of protein families in *Sistrurus barbouri* venom by cysteine mapping, N-terminal sequencing, and tandem mass spectrometry analysis. Proteomics 4:327–338

Juárez P, Wagstaff SC, Sanz L, Harrison RA, Calvete JJ (2006) Molecular cloning of *Echis ocellatus* disintegrins reveals non-venom-secreted proteins and a pathway for the evolution of ocellatusin. J Mol Evol. Companion paper DOI: 10.1007/s00239-005-0269-y

Kini R, Evans HJ (1992) Structural domains in venom proteins: evidence that metalloproteinases and nonenzymatic platelet aggregation inhibitors (disintegrins) from snake venoms are derived by proteolysis from a common precursor. Toxicon 30:265–293

Kordis D, Krizaj I, Gubensek F (2002) Functional diversification of animal toxins by adaptative evolution. In: Ménez A (ed) Perspectives in molecular toxinology. John Wiley & Sons, Chichester, UK, pp 401–419

Kumar S, Tamura K, Jakobsen IB, Nei M (2001) Mega2: Molecular Evolutionary Genetics Analysis software. Bioinformatics 17:1244–1245

Lu Q, Navdaev A, Clemetson JM, Clemetson KJ (2005a) Snake venom C-type lectins interacting with platelet receptors. Structure-function relationships and effects on hemostasis. Toxicon 45:1089–1098

Lu X, Lu D, Scully MF, Kakkar VV (2005b) Snake venom metalloproteinase containing a disintegrin-like domain, its structure-activity relationships at interacting with integrins. Curr Med Chem Cardiovasc Hematol Agents 3:249–260

Markland FS (1998) Snake venoms and the hemostatic system. Toxicon 36:1749–1800

Ménez A (2002) Perspectives in molecular toxinology John Wiley & Sons, Chichester, UK

Monleón D, Esteve V, Kovacs H, Calvete JJ, Celda B (2005) Conformation and concerted dynamics of the integrin-binding site and the C-terminal region of echistatin revealed by homonuclear NMR. Biochem J 387:57–66

Moreno-Murciano MP, Monleón D, Marcinkiewicz C, Calvete JJ, Celda B (2003) NMR solution structure of the non–RGD disintegrin obtustatin. J Mol Biol 329:135–145

Moura da Silva AM, Theakston RDG, Crampton JM (1996) Evolution of disintegrin cysteine-rich and mammalian matrix-degrading metalloproteinases: gene duplication and divergence of a common ancestor rather than convergent evolution. J Mol Evol 43:263–269

Moura da Silva AM, Marcinkiewicz C, Marcinkiewicz M, Niewiarowski S (2001) Selective recognition of α2β1 integrin by jararhagin, a metalloproteinase/disintegrin from *Bothrops jararaca* venom. Thromb Res 102:153–159

Nikai T, Taniguchi K, Komori Y, Masuda K, Fox JW, Sugihara H (2000) Primary structure and functional characterization of bilitoxin-1, a novel dimeric P-II snake venom metalloproteinase from *Agkistrodon bilineatus* venom. Arch Biochem Biophys 378:6–15

Ohno M, Ogawa T, Oda-Ueda N, Chijiwa T, Hattori S (2002) Accelerated and regional evolution of snake venom gland isozymes. In: Ménez A (ed) Perspectives in molecular toxinology. John Wiley & Sons, Chichester, UK, pp 387–401

Okuda D, Nozaki C, Sekiya F, Morita T (2001) Comparative biochemistry of disintegrins isolated from snake venoms: consideration of the taxonomy and geographical distribution of snakes in the genus *Echis*. J Biochem 129:615–620

Okuda D, Koike H, Morita T (2002) A new gene structure of the disintegrin family: a subunit of dimeric disintegrin has a short coding region. Biochemistry 41:14248–14254

Paine MJ, Desmond HP, Theakston RD, Crampton JM (1992) Purification, cloning, and molecular characterization of a high molecular weight hemorrhagic metalloprotease, jararhagin, from *Bothrops jararaca* venom. Insights into the disintegrin gene family. J Biol Chem 267:22869–22876

Park D, Kang I, Kim H, Chung K, Kim DS, Yun Y (1998) Cloning and characterization of novel disintegrins from *Agkistrodon halys* venom. Mol Cells 8:578–584

Shebuski RJ, Ramjit DR, Bencen GH, Polokoff MA (1989) Characterization and platelet inhibitory activity of bitistatin, a potent arginine-glycine-aspartic acid-containing peptide from the venom of the viper *Bitis arietans*. J Biol Chem 264:21550–21556

Shimokawa K-I, Shannon JD, Jia L-G, Fox JW (1997) Sequence and biological activity of catrocollastatin-C: a disintegrin-like/cysteine-rich two domain protein from *Crotalus atrox* venom. Arch Biochem Biophys 343:35–43

Sanz L, Bazaa A, Marrakchi N, Pérez A, Chenik M, Bel Lasfer Z, El Ayeb M, Calvete JJ (2006) Molecular cloning of disintegrins from *Cerastes vipera* and *Macrovipera lebetina transmediterranea* venom gland cDNA libraries. Insight into the evolution of the snake venom's integrin inhibition system. Biochem J 395:385–392

Souza DHF, Iemma MRC, Ferreira LL, Faria JP, Oliva MLV, Zingali RB, Niewiarowski S, de Selistre Araujo HS (2000) The disintegrin-like domain of the snake venom metalloprotease alternagin inhibits $\alpha 2\beta 1$ integrin-mediated cell adhesion. Arch Biochem Biophys 384:341–350

Tani A, Ogawa T, Nose T, Nikandrov NN, Deshimaru M, Chijiwa T, Chang CC, Fukumaki Y, Ohno M (2002) Characterization, primary structure and molecular evolution of anticoagulant protein from *Agkistrodon actus* venom. Toxicon 40:803–813

Tsai IH, Wang YM, Chiang TY, Chen YL, Huang RJ (2000) Purification, cloning and sequence analyses for pro-metalloprotease-disintegrin variants from *Deinagkistrodon acutus* venom and subclassification of the small venom metalloproteases. Eur J Biochem 267:1359–1367

Tsai I-H, Chen Y-H, Wang Y-M (2004) Comparative proteomics and subtyping of venom phospholipases A2 and disintegrins of *Protobothrops* pit vipers. Biochim Biophys Acta 1702:111–119

Zhou Q, Smith JB, Grossman MH (1995) Molecular cloning and expression of catrocollastatin, a snake-venom protein from *Crotalus atrox* (western diamondback rattlesnake) which inhibits platelet adhesion to collagen. Biochem J 307:411–417

Zigrino P, Kamiguti AS, Eble J, Drescher C, Nischt R, Fox JW, Mauch C (2002) The reprolysin jararhagin, a snake venom metalloproteinase, functions as a fibrillar collagen agonist involved in fibroblast cell adhesion and signaling. J Biol Chem 277:40528–40535

3.3. ARTÍCULO 3: Molecular cloning of *Echis ocellatus* disintegrins reveals non-venom-secreted proteins and a pathway for the evolution of Ocellatusin

J Mol Evol (2006) xx:1–11
DOI: 10.1007/s00239-005-0269-y

JOURNAL OF MOLECULAR EVOLUTION

Molecular Cloning of *Echis ocellatus* Disintegrins Reveals Non-Venom-Secreted Proteins and a Pathway for the Evolution of Ocellatusin

Paula Juárez,[1] Simon C. Wagstaff,[2] Libia Sanz,[1] Robert A. Harrison,[2] Juan J. Calvete[1]

[1] Instituto de Biomedicina de Valencia, CSIC, Jaime Roig 11, 46010 Valencia, Spain
[2] Alistair Reid Venom Research Unit, Liverpool School of Tropical Medicine, Pembroke Place, Liverpool L3 5QA, UK

Received: 10 November 2005 / Accepted: 14 March 2006 [*Reviewing Editor:* Dr. Bryan Grieg Fry]

Abstract. We report the cloning and sequence analysis of *Echis ocellatus* cDNAs coding for dimeric disintegrin subunits and for the short disintegrin ocellatusin. All the dimeric disintegrin subunit messengers belong to the short-coding class, indicating that short messengers may be more widely distributed than previously thought. Mass spectrometric analysis of the HPLC-separated venom proteins was performed to characterize the dimeric disintegrins expressed in the venom proteome. In addition to previously reported EO4 and EO5 heterodimers, a novel dimeric disintegrin containing RGD- and KGD-bearing subunits was identified. However, a WGD-containing polypeptide encoded by clone Eo1-1 was not detected in the venom, suggesting the occurrence of larger genomic than proteomic diversity, which could represent part of a non-venom-secreted reservoir of disintegrin that may eventually acquire physiological relevance for the snake upon changes of ecological niches and prey habits. On the other hand, the realization of the existence of two distinct messengers coding for the short disintegrin ocellatusin reveals key events of the evolutionary emergence of the short disintegrin ocellatusin from a short-coding dimeric disintegrin precursor by two nucleotide mutations.

Key words: Snake venom — *Echis ocellatus* — cDNA cloning — Disintegrin evolution — Ocellatusin precursor

Correspondence to: Juan J. Calvete; *email:* jcalvete@ibv.csic.es

Introduction

Approximately 130 million years ago snakes diverged from lizards and since then they developed their own venom apparatus (Kochva 1987; Fry and Wüster 2004; Fry et al. 2005). Snake venoms are composed of proteins belonging to only a few major families, including enzymes (serine proteinases, Zn^{2+}-metalloproteases, L-amino acid oxidases, phospolipases A_2 [PLA_2]) and proteins without enzymatic activity (disintegrins, C-type lectins, natriuretic [bradykinin-potentiating] peptides, cyteine-rich secretory protein [CRISP] toxins, nerve and endothelial growth factors, cystatin, and Kunitz-type protease inhibitors) (Juárez et al. 2004; Fry 2005). These toxins evolved from proteins with a normal physiological function and were recruited into the venom proteome before the diversification of the advanced snakes (Fry 2005) (superfamily Colubroidea), which make up over 80% of the about 2900 species of snake currently described (Vidal 2002), hence predating the evolution of the intricate front-fanged venom delivery mechanisms.

Venom proteins play a number of adaptative roles: immobilizing, paralyzing, killing, and digesting prey, and deterring competitors. Notably, most venom toxins are extensively cross-linked by disulfide bonds and have flourished into functionally diverse, toxin multigene families that exhibit interfamily, intergenus, interspecies, and intraspecific variability. The existence in the same venom of a functionally diverse isoform of the same protein family reflects an accelerated, by gene duplication, Darwinian evolution (Moura da Silva et al. 1996; Menez 2002; Tani et al. 2002). Rapid co-evolution between snakes and their

prey in driving the evolution of venom proteins has been discussed (Daltry et al. 1996). Studies on the underlying mechanisms of this protein structural/functional variation are of considerable theoretical interest in the field of molecular evolution, in the development of new research tools and drugs of potential clinical use, and for antivenom production strategies.

Viper venom disintegrins are a family of small (40–100 amino acids), cysteine-rich polypeptides that selectively block the function of integrin receptors (Calvete et al. 2005; Calvete 2005) and result from proteolytic processing of larger mosaic PII and PIII metalloprotease precursors (Kini and Evans 1992) or are synthesized from short-coding mRNAs (Okuda et al. 2002). Currently, the disintegrin family can be conveniently divided into five groups according to their length and number of disulfide bonds (Calvete et al. 2003). The first group includes short disintegrins, composed of 41–51 residues and 4 disulfide bonds. The second group is formed by the medium-sized disintegrins, containing about 70 amino acids and 6 disulfide bonds. The third group includes long disintegrins, with an ~84-residue polypeptide cross-linked by 7 disulfide bonds. The fourth subfamily, the PIII disintegrins, is modular proteins containing an N-terminal disintegrin-like domain of about 100 amino acids including 8 disulfide bonds and a C-terminal 110- to 120-residue cysteine-rich domain cross-linked by 6 disulfides (Calvete et al. 2002a). Unlike the PII (short, medium, and long) and PIII disintegrins, which are single-chain molecules, the fifth group is composed of homo- and heterodimers. Dimeric disintegrins contain subunits of about 67 residues with 10 cysteines involved in the formation of 4 intrachain disulfide bonds and 2 interchain cystine linkages (Calvete et al. 2000b; Bilgrami et al. 2004, 2005). Bilitoxin-1 represents a unique homodimeric disintegrin comprising disulfide-bonded polypeptides, each containing 15 cysteinyl residues (Nikai et al. 2000). The current view is that the structural diversity of disintegrins has been achieved during ophidian evolution through the selective loss of disulfide bonds (Calvete et al. 2003).

To understand the genomic basis of the accelerated evolution of disintegrins, and the molecular mechanism underlying their structural diversification, we have started the genomic analysis of cDNAs encoding disintegrins from a venom gland library of *Echis ocellatus*.

Materials and Methods

cDNA Library Synthesis and Sequencing

Total RNA was extracted (using TriReagent, Life Technologies, UK) from pooled venom glands of 10 wild-caught specimens of *E. ocellatus* (Kaltungo, Nigeria), of different ages and of both sexes, and maintained at the herpetarium of the Liverpool School of Tropical Medicine, 3 days after venom extraction when toxin gene transcription rates are at a peak (Paine et al. 1992). Messenger RNA was purified using a single-step oligo(DT) cellulose purification (Amersham, UK) according to the manufacturer's protocols. A Gateway CloneMiner cDNA library construction kit (Invitrogen, UK) was used to construct a directional plasmid cDNA library according to the manufacturer's protocol. Fractions corresponding to the first 480 ng of eluted cDNA were pooled and used for recombination and transformation into the final library. Plasmid DNA from 1000 randomly selected colonies was sequenced (Lark technologies, UK) using M13 forward primers. To identify potential toxins, Blastn or Blastx was performed on sequenced clones against databases of UniProt, translations from EMBL, and all Serpentes nucleotide sequence or protein sequences including ESTs.

cDNA Cloning and Sequencing

A forward primer, 5′-CCAAATCCAGC/TCTCCAAAATG-3′, and a reverse primer, 5′-TTCCAG/TCTCCATTGTTGG/TTTA, complementary to the highly conserved 5′- and 3′-noncoding regions of cDNA encoding for elegantin-2a from *Trimeresurus elegans* (GenBank accession number AB059572), elegantin-1a from *T. elegans* (GenBank accession number AB059571), and HR2a from *Trimeresurus flavoviridis* (accession code AY037808) were synthesized. The specific sense primer 5′-GAAAAGGAAGACG ACTGTGAATC-3′ was synthesized based on the middle portion of the clone obtained from our *E. ocellatus* EST data. To obtain the full-length cDNAs encoding these precursors, we designed an antisense primer based on the poly(A) signal: 5′-TTTTT TTTTTTTTTTTTTTTTA/C/G-3′. DNAs were amplified by PCR using total RT products and cDNA as templates. AmpliTaq Gold (Roche), a highly processive 5′–3′ DNA polymerase that lacks 3′–5′ exonuclease activity, was used. The PCR protocol included an initial denaturation step at 95°C for 6 min followed by 35 cycles of denaturation (1 min at 94°C), annealing (1 min at 55°C), and extension (1 min at 74°C) and a final extension for 7 min at 72°C. The amplified fragments were purified using the Wizard SV Gel and PCR Clean-Up System (Promega), ligated into a TA cloning vector (pCR-2.1-TOPO; Invitrogen, Groning, Netherlands), and used to transform chemically competent *Escherichia coli* (TOP 10F; Invitrogen). Positive clones, selected by growing the transformed cells in LB medium containing 10 μg/ml ampicillin, were confirmed by PCR amplification using the above primers.

Isolation and Characterization of Venom Proteins

Pooled venom extracted from 150 wild-caught *Echis ocellatus* (Kaltungo, Nigeria) of different ages and of both sexes (maintained at the herpetarium of the Liverpool School of Tropical Medicine) was lyophilized and stored at 4°C in a dark bottle. For reverse-phase HPLC separation, 2 mg of the venom was dissolved in 100 μl of 5% acetonitrile and 0.1% trifluoroacetic acid (TFA). Insoluble material was removed by centrifugation in an Eppendorf centrifuge (Hamburg, Germany) at 13,000*g* for 10 min at room temperature. Soluble proteins were separated with an ETTAN LC HPLC system (Amersham Biosciences) using a Lichrospher RP100 C18 column (250 × 4 mm, 5-mm particle size; Merck, Darmstadt, Germany) eluted at 1 ml/min with a linear gradient of 0.1% TFA in water (solution A) and in acetonitrile (solution B), first isocratically (5% B) for 5 min, followed by linear gradients of 5–45% B for 120 min and 45–70% B for 20 min. Protein detection was at 215 nm and peaks were collected manually. The isolated protein fractions were analyzed by SDS-PAGE, N-terminal sequencing (using an Applied Biosystems Procise 492 sequencer), and matrix-assisted laser-desorption ionization time-of-flight mass spectrometry (MALDI-

TOF-MS) using a Voyager-DE Pro instrument (Applied Biosystems) and sinapinic acid (3,5–dimethoxy-4-hydroxycinnamic acid; Sigma) saturated in 70% acetonitrile and 0.1% TFA as matrix, with and without prior reduction with DTT (10 mM for 15 min at 65°C) and pyridylethylation with 4-vinylpyridine (50 mM for 60 min at room temperature). The mass calibration standard consisted of a mixture of the following proteins, whose isotope-averaged molecular mass in daltons is given in parentheses: bovine insulin (5734.5), *E. coli* thioredoxin (11,674.5), and horse apomyoglobin (16,952.6). Molecular masses were also determined by electrospray ionization mass spectrometry with a triple quadrupole-ion trap hybrid instrument (QTrap from Applied Biosystems) equipped with a nanospray source (Protana, Denmark).

Phylogenetic Analysis

The program MEGA (Molecular Evolutionary Genetic Analysis; http://www.megasoftware.net (Kumar et al. 2001)) was employed for inferring phylogenies (evolutionary trees) from a multiple alignment of disintegrin sequences.

Database Accession Codes

The cDNA sequences have been deposited with the EMBL Nucleotide Sequence Data Bank (http://www.ebi.ac.uk/) and have the following accession codes: AM117387 (Dim-3 SP6), AM117388 (Eo-10), AM117389 (Eo-12), AM117390 (Eo-10c1), AM117391 (Eo1-1), and AM117392 (Eo10-c10).

Results and Discussion

Full-length disintegrin messengers were cloned from a cDNA library constructed from *Echis ocellatus* venom gland polyadenylated RNA using a forward primer for the highly conserved 5′ and either a reverse primer for the 3′ noncoding region of known disintegrins or an antisense poly(dT) primer based on the poly(A) signal. The deduced amino acid sequences of the seven cDNAs are displayed in Figs. 1 and 2.

cDNA Sequences of Dimeric Disintegrin Subunits

The deduced amino acid sequences of five *E. ocellatus* cDNAs (Fig. 1) show extensive sequence similarity to known sequences of subunits of dimeric disintegrins, including the number and position of the 10 cysteine residues (Calvete et al. 2003) (Fig. 3). Clones Dim-3 SP6 and Eo-10 (Figs. 1A and 1B, respectively) encode isoforms differing in their primary structures only at residue 54 (Fig. 3). These RGD-containing disintegrin chains are identical to EO4A (Calvete et al. 2003) except for (i) position 48, which is asparagine in EO4A and histidine in the *E. ocellatus* disintegrin cDNA, and (ii) the last four C-terminal residues, which could not be unambiguously determined by protein chemical methods and are shown here to correspond to GKYD. Except for a K/Q substitution at position 14, clone Eo-12 (Fig. 1C) corresponds to the previously reported MLD-containing A-subunit of dimeric disintegrin EO5 in *E. ocellatus* venom (Calvete et al. 2003). The deduced sequences of clones Eo-10c1 and Eo1-1 represent novel disintegrin chains. Eo-10c1 (Fig. 1D) is another isoform of Dim-3 SP6 and Eo-10 expresses a KGD integrin-inhibitory tripeptide instead of RGD.

The WGD-containing polypeptide encoded by cDNA Eo1-1 has not been reported in *Echis ocellatus* venom (Calvete et al. 2003) and may therefore correspond to a non-venom-secreted or to a novel disintegrin messenger. To check these possibilities, we reexamined the disintegrin content of the pooled venoms used for cDNA library synthesis and sequencing: venom proteins were separated by reverse-phase HPLC (Fig. 4) and each chromatographic fraction was analyzed by MALDI-TOF mass spectrometry. Dimeric disintegrin candidates were protein fractions of 13- to 15-kDa molecular mass, which generated 7- to 8-kDa subunits following reduction and pyridylethylation. The only *Echis ocellatus* venom proteins fulfilling these criteria were the previously reported EO4 and EO5 heterodimers (Calvete et al. 2003) (Fig. 4).

The molecular masses of these dimeric disintegrins were determined by nanoelectrospray ionization mass spectrometry. The mass spectrum of EO4 displayed a major (14,488 ± 2 Da) and two minor (14,583 ± 2 Da and 14,684 ± 2 Da) ions, and that of Eo5 showed a single ion of 14,520 ± 1 Da. MALDI-TOF mass analysis of the reduced and pyridylethylated (PE) proteins showed that the 14,488-Da EO4 species was composed of subunits of molecular masses of 8418 and 8196 Da. The 8196-Da subunit corresponds to the calculated molecular mass of PE-EO4B (= PE-EO5B), and the 8418-Da polypeptide has the isotope-averaged mass calculated for the cDNA-deduced PE-amino acid sequences of clone Dim-3 SP6 (8417.7 Da; = PE-EO4A). On the other hand, minor PE-EO4 ions at m/z 8293 and 8390 indicated that the 14,583-Da and the 14,684-Da EO4 species correspond to (EO4A)-S-S-(EO4B + C-terminal proline) (M_{calc}: 14,583.2 Da) and to (EO4A)-S-S-(Eo-10c1) (M_{calc}: 14,684.6 Da), respectively. The latter protein, which we call EO4-KGD to distinguish it from the previously characterized EO4-VGD species, represents a novel dimeric disintegrin. Its integrin-binding motif (KGD) has been reported to inhibit the platelet $\alpha II_b\beta_3$ integrin with a high degree of selectivity (Scarborough et al. 2001).

The isotope-averaged masses of the PE-subunits of the 14,520-Da dimeric disintegrin, 8452 and 8196 Da, confirmed the identity of this protein as EO5 (see Table 2 in Ref. 14). This protein is constituted by an 8196-Da VGD-containing B-subunit (= EO4-VGD B-subunit) and an MLD-containing A-subunit coded for by clone Eo-12 (Fig. 1C).

It is worthwhile to note that *E. ocellatus* venom glands appear to synthesize non-venom-expressed (or

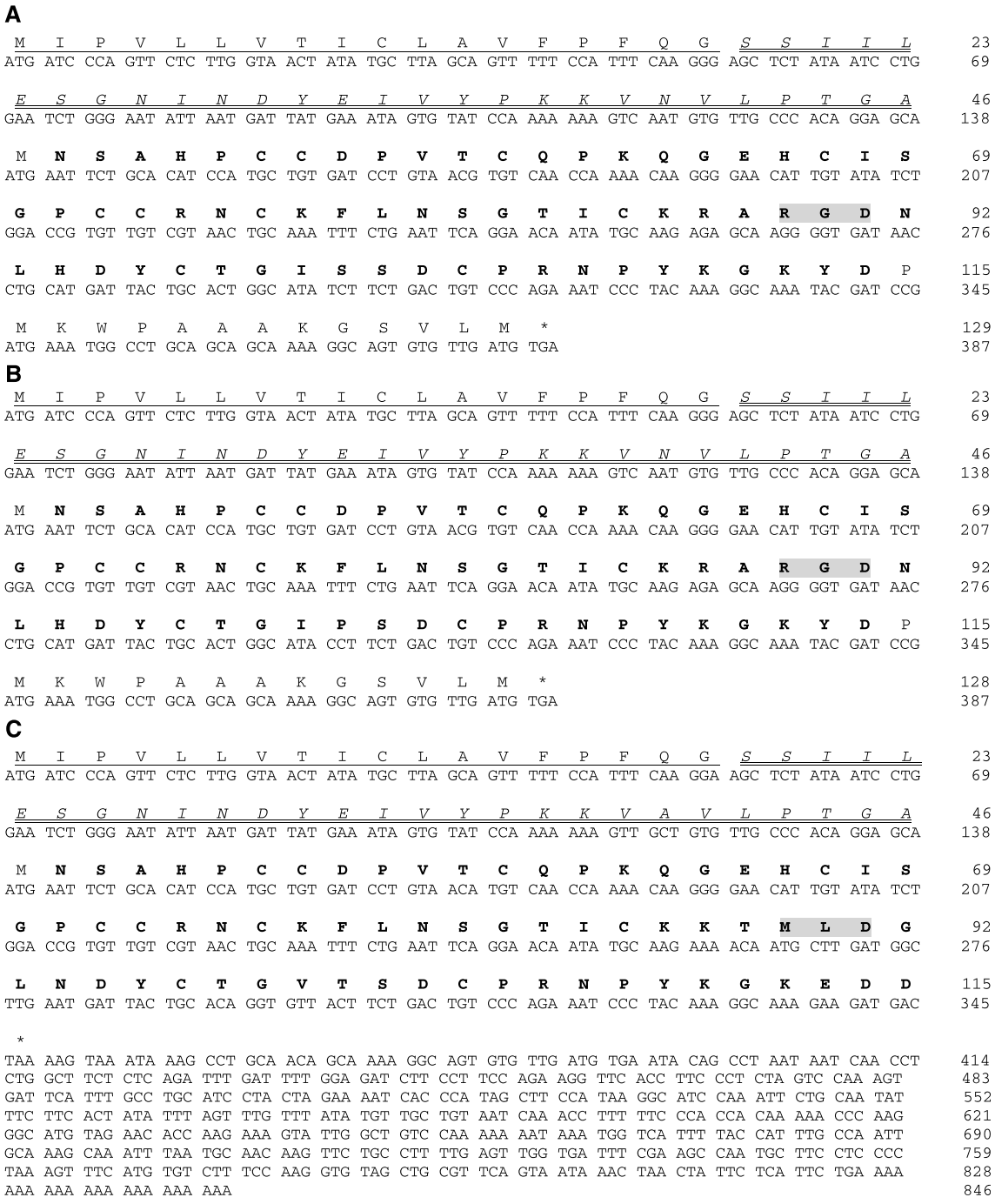


Fig. 1. DNA and deduced amino acid sequences of *Echis ocellatus* dimeric disintegrin subunits. The nucleotide sequences are numbered in the 5′–3′ direction from the initial codon ATG to the stop codon TGA or TAA (asterisks). The signal sequence and the short ened prepeptide regions are underlined and in italics and double underlined, respectively. The predicted mature protein sequences are in boldface. The integrin-binding tripeptide sequences are shaded. **A**, clone Dim-3 SP6; **B**, clone Eo-10; **C**, clone Eo-12; **D**, clone Eo-10c1; and **E**, clone Eo1-1.

very low-abundance) transcripts (Eo-10, Eo1-1) (Fig. 1) encoding putative dimeric disintegrin subunits. The WGD integrin-binding motif of Eo1-1 has been reported in studies on other viper disintegrins to enhance the inhibitory effect on the RGD-dependent integrins $\alpha_{IIb}\beta_3$, $\alpha_V\beta_3$, and $\alpha_5\beta_1$ (Calvete et al. 2002). The reason for the apparently larger transcriptomic than proteomic diversity remains unclear, although it suggests an economy of protein production. Alternatively, these messengers may exhibit an individual or a temporal expression pattern over the lifetime of the snake. We hypothesize that the non-venom-secreted or very low-abundance ($< 0.05\%$ of the total venom proteins) disintegrins may play a hitherto

Fig. 1. Continued.

unrecognized physiological function or may simply represent orphan molecules which may eventually become functional for the adaptation of snakes to changing ecological niches and prey habits. Alternatively, as pointed out in the accompanying article (Juárez et al. 2006) on the *Bitis arietans* BA-5A disintegrin, it is also possible that the undetected *Echis ocellatus* disintegrins may be translated but not secreted into the venom. Clearly, further work is needed to clarify this point.

Our results also show that short messengers coding for dimeric disintegrin subunits may be more widely distributed than previously thought, perhaps representing the canonical structure of dimeric disintegrin precursors. In support of this view, the two dimeric disintegrin messengers found in the transcriptome of *Bitis gabonica*, gabonin-1 and gabonin-2 (Francischetti et al. 2004), and all the full-length disintegrin cDNAs characterized from venom gland cDNA libraries of *Macrovipera lebetina* and *Cerastes vipera* (Sanz et al. 2006), all belong to the short-coding region class. Our working hypothesis is that removal of the metalloprotease domain may represent a step in the evolutionary pathway yielding dimeric disintegrins from a PII-metalloprotease-coding precursor (Fig. 5).

Two Distinct Messengers Coding for the Short Disintegrin Ocellatusin

Figure 2 shows the DNA and deduced amino acid sequences of a long and a short full-length *E. ocellatus* cDNA, both encoding for the short disintegrin ocellatusin (Smith et al. 2002) (DCESGP...GEHDP, isotope-averaged $M+H^{+}_{calc}$, 5594.3; $M+H^{+}_{measured}$, 5594), the major secreted disintegrin in *E. ocellatus* venom (Fig. 4). The amino acid sequences located N-terminal of mature ocellatusin within the long and the short precursors, ELLQNSVNPCYDPVTCQPKEKE and NSAHPCYDPVTCQPKEKE, respectively (Figs. 2 and 3B), exhibit large similarity between themselves and with the N-terminal regions of the subunits of dimeric disintegrins (Fig. 3A). This fact strongly argues for a common ancestry of both ocellatusin messengers and the precursors of dimeric disintegrin chains. This hypothesis is supported by a phylogenetic analysis of the dimeric and the short disintegrins from *Echis* species (Fig. 6A), which shows that the two ocellatusin precursors, EO10c-10 and EO-00006, cluster into an intermediate group between the venom-secreted dimeric disintegrin subunits and the short disintegrins. Phylogeny reconstruction of the

A

```
|----- Signal sequence ----->                                                  |-- Pro-domain -->
 M   I   Q   V   L   L   V   T   I   C   L   A   V   F   P   F   Q   G   S   S   K   T   L      23
ATG ATC CAA GTT CTC TTG GTA ACT ATA TGC TTA GCA GTT TTT CCA TTT CAA GGG AGC TCT AAA ACC CTG      69

 K   S   G   N   V   N   D   Y   E   V   V   N   P   Q   K   I   T   G   L   P   V   G   A      46
AAA TCT GGG AAT GTT AAT GAT TAT GAA GTA GTG AAT CCA CAA AAA ATC ACT GGG TTG CCT GTA GGA GCT     138

 F   K   Q   P   E   K   K   Y   E   D   A   V   Q   Y   E   F   E   V   N   G   E   P   V      69
TTT AAG CAG CCA GAG AAA AAG TAT GAA GAC GCT GTG CAA TAT GAA TTT GAA GTG AAT GGA GAG CCA GTG     207

 I   L   H   L   E   K   N   K   G   L   F   S   E   D   Y   S   E   T   H   Y   S   P   D      92
ATC CTT CAT CTG GAA AAA AAT AAA GGA CTT TTT TCA GAA GAT TAC AGT GAG ACT CAT TAT TCC CCT GAT     276

 G   S   E   I   T   T   N   P   P   V   E   D   H   C   Y   Y   H   G   H   V   Q   N   D     115
GGC AGC GAA ATT ACA ACA AAC CCT CCT GTT GAG GAT CAC TGC TAT TAT CAT GGA CAC GTC CAG AAT GAT     345

 A   D   S   T   A   S   I   S   T   C   N   G   L   K   G   F   F   T   L   R   G   E   T     138
GCT GAC TCA ACT GCA AGC ATC AGC ACA TGC AAT GGT TTG AAA GGA TTT TTT ACG CTT CGT GGG GAG ACG     414

 Y   L   I   E   P   L   K   V   P   D   S   E   S   H   A   V   Y   K   Y   E   D   A   K     161
TAC TTA ATT GAA CCC TTG AAG GTT CCC GAC AGT GAA TCC CAT GCA GTC TAC AAA TAT GAA GAT GCC AAA     483

 K   K   D   E   A   P   K   M   C   G   V   T   L   T   N   W   E   S   D   E   P   I   K     184
AAA AAG GAT GAG GCC CCC AAA ATG TGT GGG GTA ACC CTG ACT AAT TGG GAA TCA GAT GAG CCC ATC AAA     552

                            |----- Metalloproteinase ----->
 K   A   S   H   L   V   A   T   S   E   Q   Q   H   F   H   P   R   Y   V   Q   L   V   I     207
AAG GCT TCT CAT TTA GTT GCT ACT TCT GAA CAA CAG CAT TTT CAC CCA AGA TAC GTT CAG CTT GTC ATA     621

 V   A   D   H   S   M   V   T   K   N   N   N   D   L   T   A   L   T   T   W   I   H   Q     230
GTT GCA GAC CAC TCA ATG GTC ACG AAA AAC AAC AAT GAT TTA ACT GCT TTA ACA ACA TGG ATA CAT CAA     690

 I   V   N   D   M   I   V   M   Y   R   I   L   N   I   H   I   T   L   A   N   V   X   I     253
ATT GTC AAC GAT ATG ATT GTG ATG TAC AGA ATT CTG AAT ATT CAT ATA ACA CTG GCT AAC GTA NAA ATT     759

 W   S   S   G   D   L   I   T   V   T   S   S   A   P   T   T   L   R   S   F   G   E   W     276
TGG TCC AGT GGA GAT TTG ATT ACT GTG ACA TCA TCA GCA CCT ACT ACT TTG AGG TCA TTT GGA GAA TGG     828

 R   A   R   N   L   V   N   R   I   T   H   D   N   A   Q   L   I   T   A   V   H   L   D     299
AGA GCG AGA AAT TTG GTG AAT CGC ATA ACG CAT GAT AAT GCT CAA TTA ATC ACA GCC GTT CAC CTT GAT     897

 N   L   I   G   Y   G   Y   L   G   T   M   C   D   P   Q   S   S   V   A   I   T   E   D     322
AAC CTT ATA GGA TAC GGT TAC TTA GGT ACT ATG TGC GAT CCA CAA TCG TCT GTA GCA ATT ACT GAG GAT     966

 H   S   T   D   H   L   W   V   A   A   T   M   A   H   E   M   G   H   N   L   G   M   N     345
CAT AGC ACA GAT CAT CTT TGG GTT GCA GCT ACA ATG GCC CAT GAG ATG GGT CAT AAT CTG GGT ATG AAT    1035

 H   D   G   N   Q   C   N   C   G   A   A   G   C   I   M   S   A   I   I   S   Q   Y   R     368
CAT GAT GGA AAT CAG TGT AAT TGT GGT GCT GCC GGA TGC ATT ATG TCT GCG ATC ATA TCA CAA TAC CGT    1104

 S   Y   Q   F   S   D   C   S   M   N   E   Y   R   N   Y   I   T   T   H   N   P   P   C     391
TCC TAT CAG TTC AGT GAT TGT AGT ATG AAT GAA TAT CGC AAC TAT ATT ACT ACT CAT AAC CCA CCA TGC    1173

                                                                        |-- Disintegrin -->
 I   L   N   Q   A   L   R   T   D   T   V   S   T   P   V   S   E   N   E   L   L   Q   N     414
ATT CTC AAT CAA GCC CTG AGA ACA GAT ACT GTT TCA ACT CCA GTT TCT GAA AAT GAA CTT TTG CAG AAT    1242

 S   V   N   P   C   Y   D   P   V   T   C   Q   P   K   E   K   E   D   C   E   S   G   P     437
TCT GTA AAT CCA TGC TAT GAT CCT GTA ACA TGT CAA CCA AAA GAA AAG GAA GAC TGT GAA TCT GGA CCA    1311

 C   C   D   N   C   K   F   L   K   E   G   T   I   C   K   M   A   R   G   D   N   M   H     460
TGT TGT GAT AAC TGC AAA TTT CTG AAG GAA GGA ACA ATA TGC AAG ATG GCA AGG GGT GAT AAC ATG CAT    1380

 D   Y   C   N   G   K   T   C   D   C   P   R   N   P   Y   K   G   E   H   D   P   M   E     483
GAT TAC TGC AAT GGC AAA ACT TGT GAC TGT CCC AGA AAT CCT TAC AAA GGC GAA CAT GAT CCG ATG GAA    1449

 W   P   A   P   A   K   G   S   V   L   M   *                                                 495
TGG CCT GCA CCA GCA AAA GGC AGT GTG TTG ATG TGA                                               1485
```

Fig. 2. cDNA and deduced amino acid sequence of ocellatusin long (clone Eo-00006; **A**) and short (clone Eo10c-10; **B**) precursors. The nucleotide sequences are numbered in the 5′–3′ direction, from the initial codon ATG to the stop codon TGA (asterisk). The N-terminal regions of the signal sequence, the pro-domain, the metalloproteinase, and the disintegrin domains are labeled. The signal sequences are underlined. The positions of the Cys-switch site (KMCGV) within the pro-domain, the Zn^{+2}-binding motif (HEMGHNLGMNH) of the metalloprotease domain, and the RGD sequence of the disintegrin domain are shaded. The short-ened propeptide region of the short-coding ocellatusin precursor and its homologous region in the long precursor are in italics and double underlined. The short-disintegrin ocellatusin is indicated in boldface. The tyrosine residue of Eo-00006 and Eo-10c1, which in dimeric disintegrin subunits corresponds to a conserved cysteine involved in an intersubunit disulfide, is in boldface, underlined, and boxed. The short disintegrin-specific cysteine residues at positions 468 and 102 of the long and the short precursors, respectively, are double underlined.

B

```
 M   I   P   V   L   L   V   T   I   C   L   A   V   F   P   F   Q   G   S   S   I   I   L      23
ATG ATC CCA GTT CTC TTG GTA ACT ATA TGC TTA GCA GTT TTT CCA TTT CAA GGG AGC TCT ATA ATC CTG      69

 E   S   G   N   I   N   D   Y   E   I   V   Y   P   K   K   V   A   V   L   P   T   G   A      46
GAA TCT GGG AAT ATT AAT GAT TAT GAA ATA GTG TAT CCA AAA AAA GTT GCT GTG TTG CCC ACA GGA GCA     138

|---- Disintegrin ----->
 M   N   S   A   H   P   C   Y   D   P   V   T   C   Q   P   K   E   K   E   D   C   E   S      69
ATG AAT TCT GCA CAT CCA TGC TAT GAT CCT GTA ACA TGT CAA CCA AAA GAA AAG GAA GAC TGT GAA TCT     207

 G   P   C   C   D   N   C   K   F   L   K   E   G   T   I   C   K   M   A   R   G   D   N      92
GGA CCA TGT TGT GAT AAC TGC AAA TTT CTG AAG GAA GGA ACA ATA TGC AAG ATG GCA AGG GGT GAT AAC     276

 M   H   D   Y   C   N   G   K   T   C   D   C   P   R   N   P   Y   K   G   E   H   D   P     115
ATG CAT GAT TAC TGC AAT GGC AAA ACT TGT GAC TGT CCC AGA AAT CCT TAC AAA GGC GAA CAT GAT CCG     345

 M   E   W   P   A   P   A   K   G   S   V   L   M   *                                         128
ATG GAA TGG CCT GCA CCA GCA AAA GGC AGT GTG TTG ATG TGA AAA CAG CCT ACT AAT CAA CCT CTG GCT     414
TCT CTC AGA TTT GAT TTT GGA GAT CCT TCT TCC TGA AGG TTC ACC TTC CCT CTA GTC CAA AGA GAT TCA     483
TTT GCC TGC ATC CTA CTA GTA AAT CAC CCT TAG CTT CCA GAT GGC ATT CAA ATT CTG CAA TAT TTC TTC     552
ACT ATA TTT AGT TTG TTT ACA TTC TGC TGT AAT CAA ACC TTT TTC CCA CCA CAA AAA TCC AAG GGC ATG     621
TAG AAC ACC AAG AAA GTA TTG GCT GTC CAA AAA AAT AAA TGG TCA TTT TAC CAT TTG CCA ATT GCA AAG     690
CAA ATT TAA TGC AAC AAG TTC TGC CTT TTG AGT TGG TGA TTT CGA AGC CAA TGC TTC CTC CCC TAA AGT     759
TTC ATG TGC CTT TCC AAA GTG TAG CTG CGT TCA ATA ATA AAC TAA CTA TTC TCA TTC TGA AAA AAA AAA     828
AAA AAA AAA AAA A                                                                           841
```

Fig. 2b. Continued.

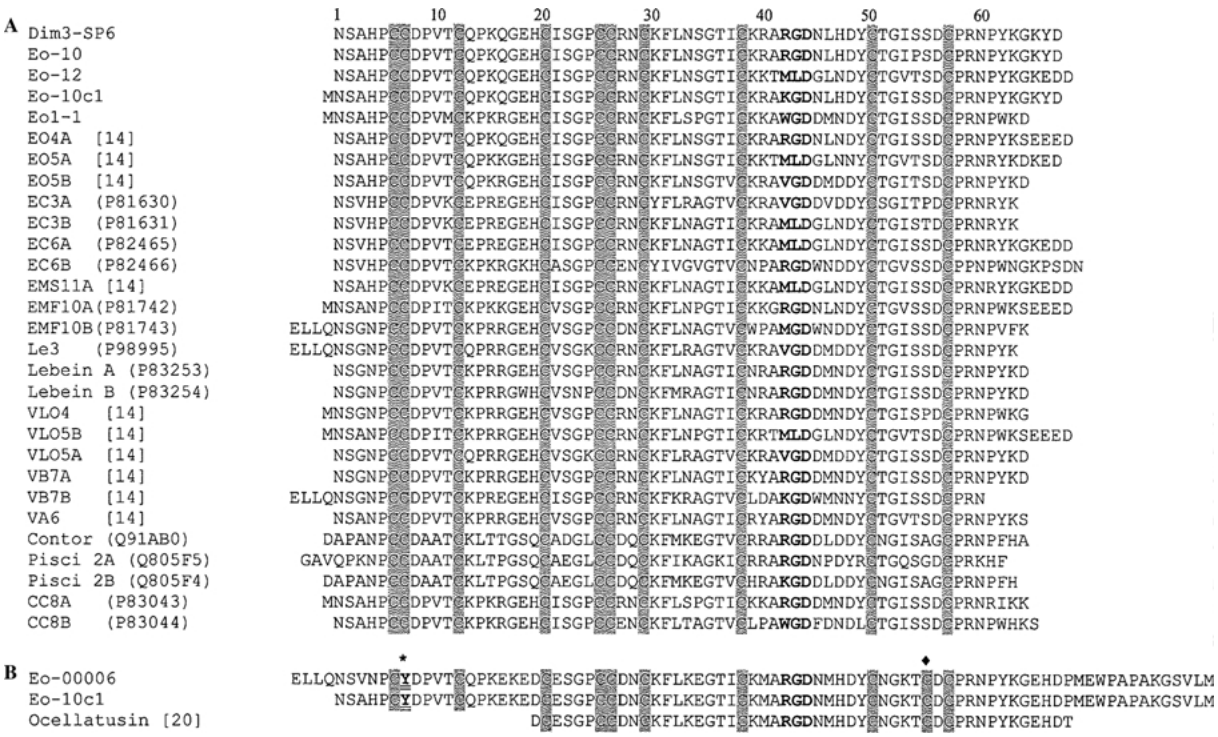


Fig. 3. **A** Multiple amino acid sequence alignment of the *Echis ocellatus* cDNA-deduced disintegrin sequences with those of dimeric disintegrin subunits. **B** Alignment of the cDNA-deduced amino acid sequences of the long (Eo-00006) and short (Eo10c-10) ocellatusin precursors with that of the protein isolated from the venom of *Echis ocellatus* (Smith et al. 2002). Cysteine residues are shaded in pale gray. The integrin-binding tripeptides are shown in boldface. The tyrosine residue of Eo-00006 and Eo-10c1, which in dimeric disintegrin subunits corresponds to a conserved cysteine involved in an intersubunit disulfide, is highlighted in boldface, double underlined, and marked with an asterisk. The short disintegrin-specific cysteine residue is labeled by a filled diamond.When available, the Swiss-Prot/TrEMBL (http://us.expasy.org/sprot/) databank accession codes are given in brackets.

known disintegrins of *Echis ocellatus* using a maximum parsimony method (Fig. 6B) also supports the emergence of ocellatusin from EO10c-10 and EO-00006 transcripts.

Snake venom PIII disintegrins evolved from the extracellular domains of cell membrane-anchored ADAM (a disintegrin and metalloprotease) molecules after mammals and reptiles diverged (Moura

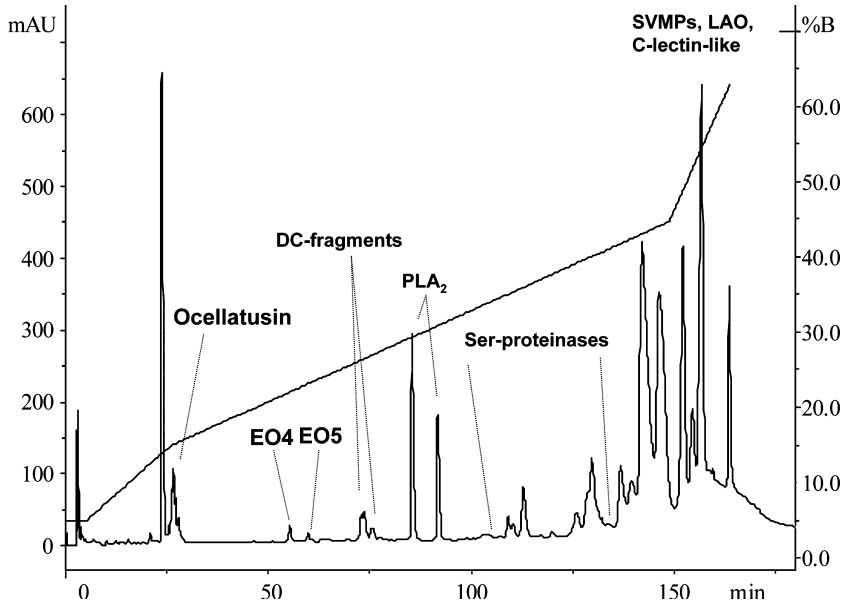


Fig. 4. Reverse-phase HPLC separation of *Echis ocellatus* venom proteome. Protein fractions whose identities were characterized by combination of N-terminal sequencing and mass spectrometry are labeled. Peaks containing the short disintegrin ocellatusin and the dimeric disintegrins EO4 and EO5 are highlighted in boldface. DC-fragments, disintegrin/cysteine-rich domains of PIII metalloproteases; PLA_2, phospholipase A2; SVMPs, snake venom metalloproteases; LAO, L-amino acid oxidase.

da Silva et al. 1996). Phylogenetic analysis, in conjunction with biochemical and genetic data, supports the model depicted in Fig. 6C, by which the structural diversification of the disintegrin family occurred through the successive loss of disulfide bonds (Calvete et al. 2003).

The most noticeable difference between ocellatusin and the dimeric disintegrin precursors is the presence at position 54 of the ocellatusin short-coding precursor (position 420 of the long precursor) of a tyrosine residue, which in all dimeric disintegrin subunits harbors a conserved cysteine residue (Figs. 2 and 3) involved in dimer formation (Calvete et al. 2000b; Bilgrami et al. 2004, 2005). In crystallographic studies, this Cys7 residue of each subunit of two *E. carinatus sochureki* dimeric disintegrins forms the two interchain cystine linkages (Cys7A-Cys12B and Cys12A-Cys7B) with cysteine-12 from the other subunit of the dimers (Bilgrami et al. 2004, 2005) (Fig. 5). Theoretically, the single amino acid substitution Cys7 → Tyr, generated through a single nucleotide mutation, T<u>G</u>T → T<u>A</u>T or T<u>G</u>C → T<u>A</u>C, is the simplest molecular solution to hinder dimerization. Though experimental validation is needed, we hypothesize that the Cys7 → Xaa mutation may also represent a key step along the evolutionary pathway of short disintegrins from dimeric disintegrin precursors. Substitution, by cysteine, of the residue located two positions N-terminal of the last cysteine residue of the disintegrin domain (residues 468 and 102 of the long and the short precursors, respectively) (Fig. 2) and subsequent proteolytic processing of the N- and the C-terminal regions are further molecular events needed for generating mature short disintegrins (Fig. 5). In this sense, it is noteworthy that in all cDNAs coding for short dimeric disintegrin messengers, residue 102 is either a serine invariably coded for by a TCT co-

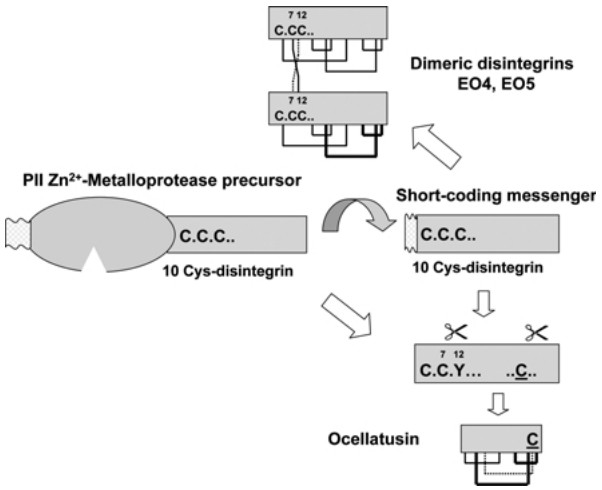


Fig. 5. Illustration depicting the proposed common ancestry of the messenger precursors coding for the short disintegrin ocellatusin and dimeric disintegrins EO4 and EO5. The proposed evolutionary pathway of EO4 and EO5 includes the removal of the metalloproteinase domain from a PII metalloproteinase precursor gene. A key event in the emergence of ocellatusin appears to be the substitution of the second N-terminal cysteine residue (Cys7 in the dimeric disintegrin subunit precursor) by tyrosine, thereby impairing dimerization through homologous CysA7-CysB12 and CysA12-CysB7 linkages, along with the appearance of a novel cysteine residue at position 102 (short-coding precursor numbering) between the 9th and the 10th cysteine of the precursor (<u>C</u>), enabling the short disintegrin-specific disulfide bond depicted by the dashed line, and the proteolytic processing of the N- and the C-terminal regions (scissors). The two conserved disulfide bonds in the structures of dimeric disintegrin subunits and the short disintegrins are represented by thick lines.

don, including acostatin α from *Agkistrodon contortrix contortrix* (Okuda et al. 2002), clones Dim-3 SP6, Eo-10, Eo-12, Eo-10c1, and Eo1-1 from *Echis ocellatus* (Fig. 1), clones CV3 and CV11 from *Cerastes vipera*, and ML2/8/15 and ML3 from *Macrovipera lebetina* (Sanz et al. 2006), or a threonine (ACT codon) in gabonin-1 and gabonin-2 from *Bitis gabonica*

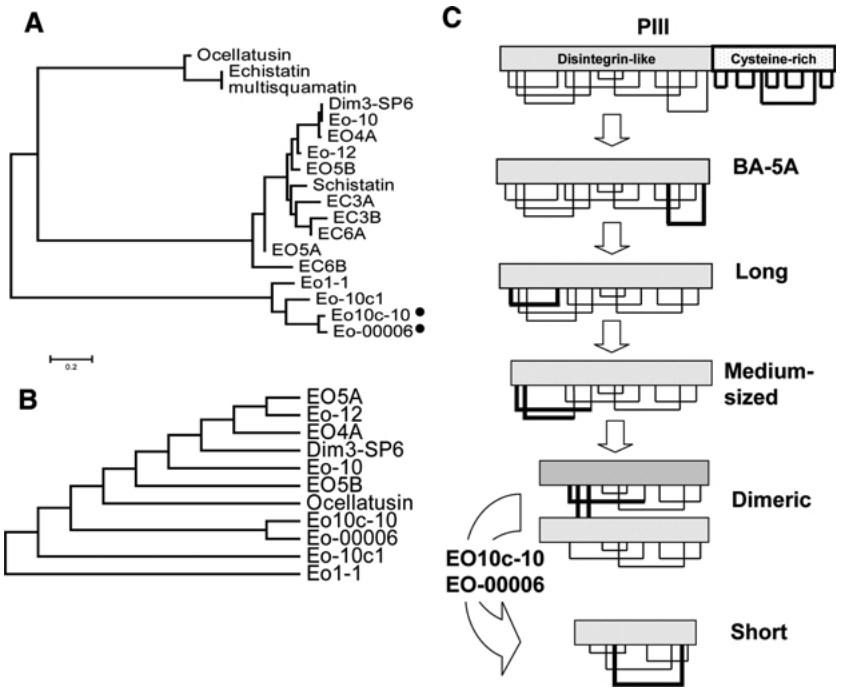


Fig. 6. **A** Cladogram for the multiple sequence analysis of the dimeric disintegrin subunits and the short disintegrins from *Echis* snake species. The ocellatusin precursors, EO10c-10 and EO-00006, are denoted by filled circles. The tree represents the minimum evolutionary distance estimated through neighbor joining using maximum likelihood distances. The length of the horizontal scale bar represents 20% divergence. For primary references of the analyzed disintegrins, consult Calvete et al. (2003). Maximum parsimony analysis produced a similar topology (**B**). **C** Scheme of the domain organization, disulfide bond patterns, and proposed evolutionary pathway from the PIII disintegrin/cysteine-rich proteins to short disintegrins. Structural features (the cysteine-rich domain of PIII disintegrin-like molecules, and class-specific disulfides) lost along the disintegrin diversification pathway are highlighted with thick lines. The proposed mechanism for the emergence of the short disintegrin ocellatusin from the long (EO-00006) and the short (EO10c-10) precursors is shown schematically in Fig. 5.

(Francischetti et al. 2004). Substitution of serine for cysteine can be accomplished by a single C → G mutation, whereas changing threonine for cysteine needs a minimum of two mutations (i.e., ACT → TCT → TGT). This may in part explain the observation that short disintegrins are commonly found in the venoms of species from genera like *Echis*, which also express dimeric disintegrins containing the sequence "**SXDC**" but have not been reported in the venom gland transcriptome of *Bitis gabonica* (dimeric disintegrins with "**TPDC**" sequence).

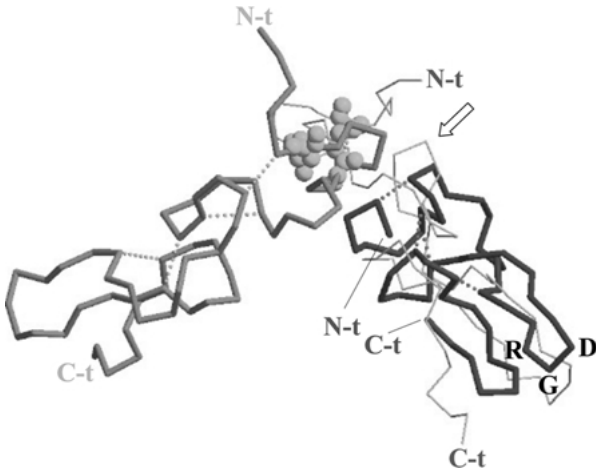


Fig. 7. Superposition of Cα traces of echistatin (red) and schistatin (magenta and cyan). N-t, N-terminus; C-t, C-terminus. Intrachain disulfide bonds are shown in one subunit of schistatin as dashed green lines, and the two interchain cystine linkages of the homodimer are depicted in the space-filling model in green. Disulfide bonds of echistatin are shown as dashed orange lines, and the RGD motif of the short disintegrin is labeled. The arrow indicates the position of the N-terminal proteolytic cleavage (E^{65}-D^{66} of the short-coding precursor shown in Fig. 2B) depicted as a scissor in Fig. 5.

Concluding Remarks

The specific goal of this article was to investigate the genomic basis of the accelerated evolution of disintegrins and the molecular mechanism underlying their structural diversification. To this end, we sought to analyze cDNAs encoding disintegrins from a venom gland library of *Echis ocellatus*. The combined transcriptomic and proteomic analysis of *E. ocellatus* disintegrins revealed a non-venom-secreted disintegrin, suggesting that the accepted view of an accelerated evolution within the disintegrin family might not be mirrored in the venom proteome. A "hidden reservoir" of non-venom-secreted toxins could represent a pool of orphan molecules which may eventually become of relevance, i.e., for the adaptation of snakes in response to a changing ecological niche and prey habits. The identification of two distinct messengers coding for the short disintegrin ocellatusin sheds light on key events of the evolutionary pathway leading to short-coding messengers for dimeric disintegrin subunits from a PII-metalloproteinase precursor and to the short disintegrin ocellatusin from a short-coding dimeric disintegrin precursor by a minimum of two nucleotide mutations (Cys → Tyr and Ser → Cys). It is notable that the native fold of short disintegrins adopts a disulfide bond pattern slightly different from that of the dimeric disintegrin chains (Calvete et al. 2003; Moreno-Murciano et al. 2003; Monleón et al. 2005) (Fig. 5), providing additional possibilities for the evolution of the structure and

function of this family of integrin antagonists. This point is illustrated in Fig. 7, in which the three-dimensional structures of schistatin (homodimer; PDB code 1RMR) (Bilgrami et al. 2004) and echistatin (short disintegrin; PDB code 1RO3) (Monleón et al. 2005), both from *E. carinatus sochureki*, have been superimposed. The root mean square deviation (r.m.s.d.) for the topologically equivalent Cα atoms is 4–5 Å, and the loop that contains the RGD motif and the C-terminal tail, which exhibit concerted dynamics, are oriented differently in the two structures. However, the overall scaffolding of schistatin has been reported to display an r.m.s.d. of 1.6 Å for the Cα atoms with medium-sized trimestatin (Bilgrami et al. 2004) reflecting, at the structural level, the proposed divergent evolution of disintegrins by disulfide bond engineering (Calvete et al. 2003).

Acknowledgments. This work was financed by Grant BFU2004-01432 from the Ministerio de Educación y Ciencia, Madrid, Spain (to J.J.C.). P.J. and L.S. are recipients of a predoctoral fellowship (FPI; formación de personal investigador) from the Spanish Ministerio de Educación y Ciencia and a postdoctoral I3P contract, respectively. R.A.H. and S.C.W. were funded by the Wellcome Trust.

References

Bilgrami S, Tomar S, Yadav S, Kaur P, Kumar J, Jabeen T, Sharma S, Sinhg TP (2004) Crystal structure of schistatin, a disintegrin homodimer from saw-scaled viper (*Echis carinatus*) at 2.5 Å resolution. J Mol Biol 341:829–837

Bilgrami S, Yadav S, Sharma S, Perbandt M, Betzel C, Singh TP (2005) Crystal structure of the disintegrin heterodimer from saw-scaled viper (*Echis carinatus*) at 1.9 Å resolution. Biochemistry 44:11058–11066

Calvete JJ (2005) Structure-function correlations of snake venom disintegrins. Curr Pharm Des 11:829–835

Calvete JJ, Moreno-Murciano MP, Sanz L, Jürgens M, Schrader M, Raida M, Benjamin DC, Fox JW (2000a) The disulfide bond pattern of catrocollastatin C, a disintegrin/cysteine-rich protein isolated from *Crotalus atrox* venom. Protein Sci 9:1365–1373

Calvete JJ, Jürgens M, Marcinkiewicz C, Romero A, Schrader M, Niewiarowski S (2000b) Disulfide bond pattern and molecular modelling of the dimeric disintegrin EMF-10, a potent and selective integrin α5β1 antagonist from *Eristocophis macmahoni* venom. Biochem J 345:573–581

Calvete JJ, Fox JW, Agelan A, Niewiarowski S, Marcinkiewicz C (2002) The presence of the WGD motif in CC8 heterodimeric disintegrin increases its inhibitory effect on αIIbβ3, αVβ3, and α5β1 integrins. Biochemistry 41:2014–2021

Calvete JJ, Moreno-Murciano MP, Theakston RDG, Kisiel DG, Marcinkiewicz C (2003) Snake venom disintegrins: novel dimeric disintegrins and structural diversification by disulfide bond engineering. Biochem J 372:725–734

Calvete JJ, Marcinkiewicz C, Monleón D, Esteve V, Celda B, Juárez P, Sanz L (2005) Snake venom disintegrins: evolution of structure and function. Toxicon 45:1063–1074

Daltry JC, Wüster W, Thorpe RS (1996) Diet and snake venom evolution. Nature 379:537–540

Francischetti IMB, My-Pham V, Harrison J, Garfield MK, Ribeiro JMC (2004) *Bitis gabonica* (Gaboon viper) snake venom gland: toward a catalog for the full-length transcripts (cDNA) and proteins. Gene 337:55–69

Fry BG (2005) From genome to "venome": molecular origin and evolution of the snake venom proteome inferred from phylogenetic analysis of toxin sequences and related body proteins. Genome Res 15:403–420

Fry BG, Wüster W (2004) Assembling an arsenal: origin and evolution of the snake venom proteome inferred from phylogenetic analysis of toxin sequences. Mol Biol Evol 21:870–883

Fry BG, Vidal N, Norman JA, Vonk FJ, Scheib H, Ramjan SFR, Kuruppu S, Fung K, Hedges SB, Richardson MK, Hodgson WC, Ignjatovic V, Summerhayes R, Kochva E (2005) Early evolution of the venom system in lizards and snakes. Nature 439:584–588

Juárez P, Sanz L, Calvete JJ (2004) Snake venomics: characterization of protein families in *Sistrurus barbouri* venom by cysteine mapping, N-terminal sequencing, and tandem mass spectrometry analysis. Proteomics 4:327–338

Juárez P, Wagstaff SC, Oliver J, Sanz L, Harrison RA, Calvete JJ (2006) Molecular cloning of disintegrin-like transcript BA-5A from a *Bitis arietans* venom gland cDNA library: a putative intermediate in the evolution of the long-chain disintegrin bitistatin. J Mol Evol. Companion paper 10.1007/s00239-006-0268-z

Kini R, Evans HJ (1992) Structural domains in venom proteins: evidence that metalloproteinases and nonenzymatic platelet aggregation inhibitors (disintegrins) from snake venoms are derived by proteolysis from a common precursor. Toxicon 30:265–293

Kochva E (1987) The origin of snakes and evolution of the venom apparatus. Toxicon 25:65–106

Kumar S, Tamura K, Jakobsen IB, Nei M (2001) Mega2: Molecular Evolutionary Genetics Analysis software. Bioinformatics 17:1244–1245

Ménez A (2002) Perspectives in molecular toxinology. John Wiley & Sons, Chichester, UK

Monleón D, Esteve V, Kovacs H, Calvete JJ, Celda B (2005) Conformation and concerted dynamics of the integrin-binding site and the C-terminal region of echistatin revealed by homonuclear NMR. Biochem J 387:57–66

Moreno-Murciano MP, Monleón D, Marcinkiewicz C, Calvete JJ, Celda B (2003) NMR solution structure of the non-RGD disintegrin obtustatin. J Mol Biol 329:135–145

Moura Da Silva AM, Theakston RDG, Crampton JM (1996) Evolution of disintegrin cysteine-rich and mammalian matrix-degrading metalloproteinases: gene duplication and divergence of a common ancestor rather than convergent evolution. J Mol Evol 43:263–269

Nikai T, Taniguchi K, Komori Y, Masuda K, Fox JW, Sugihara H (2000) Primary structure and functional characterization of bilitoxin-1, a novel dimeric P-II snake venom metalloproteinase from *Agkistrodon bilineatus* venom. Arch Biochem Biophys 378:6–15

Okuda D, Koike H, Morita T (2002) A new gene structure of the disintegrin family: a subunit of dimeric disintegrin has a short coding region. Biochemistry 41:14248–14254

Paine MJ, Desmond HP, Theakston RD, Crampton JM (1992) Purification, cloning, and molecular characterization of a high molecular weight hemorrhagic metalloprotease, jararhagin, from *Bothrops jararaca* venom. Insights into the disintegrin gene family. J Biol Chem 267:22869–22876

Sanz L, Bazaa A, Marrakchi N, Pérez A, Chenik M, Bel Lasfer Z, El Ayeb M, Calvete JJ (2006) Molecular cloning of disintegrins from *Cerastes vipera* and *Macrovipera lebetina transmediterranea* venom gland cDNA libraries. Insight into the evolution of

the snake venom's integrin inhibition system. Biochem J 395:385–392

Smith JB, Theakston RDG, Coelho ALJ, Barja-Fidalgo C, Calvete JJ, Marcinkiewicz C (2002) Characterization of a monomeric disintegrin, ocellatusin, present in the venom of the Nigerian carpet viper, *Echis ocellatus*. FEBS Lett 512:111–115

Scarborough RM, Rose JW, Hsu MA, Phillips DR, Fried VA, Campbell AM, Nannizzi L, Charo IF (2001) Barbourin, a GPIIb-IIIa-specific integrin antagonist from the venom of *Sistrurus m. barbouri*. J Biol Chem 266:9359–9362

Tani A, Ogawa T, Nose T, Nikandrov NN, Deshimaru M, Chijiwa T, Chang CC, Fukumaki Y, Ohno M (2002) Characterization, primary structure and molecular evolution of anticoagulant protein from *Agkistrodon actus* venom. Toxicon 40:803–813

Vidal N (2002) Colubroid systematics: evidence for an early appearance of the venom apparatus followed by extensive evolutionary tinkering. J Toxicol Toxin Rev 21:21–41

3.4. ARTÍCULO 4: Loss of introns along the evolutionary diversification pathway of snake venom disintegrins evidenced by sequence analysis of genomic DNA from *Macrovipera lebetina transmediterranea* and *Echis ocellatus*

J Mol Evol (2006) xx:1–11
DOI: 10.1007/s00239-006-0161-4

JOURNAL OF MOLECULAR EVOLUTION

Loss of Introns Along the Evolutionary Diversification Pathway of Snake Venom Disintegrins Evidenced by Sequence Analysis of Genomic DNA from *Macrovipera lebetina transmediterranea* and *Echis ocellatus*

Amine Bazaa,[1,*] Paula Juárez,[2,*] Néziha Marrakchi,[1] Zakaria Bel Lasfer,[1] Mohamed El Ayeb,[1] Robert A. Harrison,[3] Juan J. Calvete,[2] Libia Sanz[2]

[1] Laboratoire des Venins et Toxines, Institut Pasteur de Tunis, B.P. 74, 1002 Tunis-Belvédère, Tunisia
[2] Instituto de Biomedicina de Valencia, CSIC, Jaime Roig 11, 46010 Valencia, Spain
[3] Alistair Reid Venom Research Unit, Liverpool School of Tropical Medicine, Pembroke Place Liverpool L3 5QA, UK

Received: 13 July 2006 / Accepted: 9 October 2006 [*Reviewing Editor:* Dr. Bryan Fry]

Abstract. Analysis of cDNAs from *Macrovipera lebetina transmediterranea* (Mlt) and *Echis ocellatus* (Eo) venom gland libraries encoding disintegrins argued strongly for a common ancestry of the messengers of short disintegrins and those for precursors of dimeric disintegrin chains. We now report the sequence analysis of disintegrin-coding genes from these two vipers. Genomic DNAs for dimeric disintegrin subunits Ml_G1 and Ml_G2 (Mlt) and Eo_D3 (Eo) contain single 1-kb introns exhibiting the 5′-GTAAG (donor)/3′-AG (acceptor) consensus intron splicing signature. On the other hand, the short RTS-disintegrins Ml_G3 (Mlt) and Eo_RTS (Eo) and the short RGD-disintegrin ocellatusin (Eo) are transcribed from intronless genomic DNA sequences, indicating that the evolutionary pathway leading to the emergence of short disintegrins involved the removal of all intronic sequences. The insertion position of the intron within Ml_G1, Ml_G2, and Eo_D3 is conserved in the genes for vertebrate ADAM (A disintegrin and metalloproteinase) protein disintegrin-like domains and within the gene for the medium-size snake disintegrins halystatins 2 and 3. However, a comparative analysis of currently available disintegrin(-like) genes outlines the view that a minimization of both the gene organization and the protein structure underlies the evolution of the snake venom disintegrin family.

Key words: *Macrovipera lebetina transmediterranea* — *Echis ocellatus* — Genomic DNA — Intron sequence — Intronless gene — Disintegrin evolution

Introduction

Venom represents a key innovation in ophidian evolution that allowed advanced snakes to transition from a mechanical (constriction) to a chemical (venom) means of subduing and digesting prey larger than themselves. Venom toxins likely evolved from endogenous proteins with normal physiological functions that were recruited into the venom proteome before the radiation of the advanced snakes at the base of the Colubroidea radiation (Fry and Wüster 2004; Fry 2005; Fry et al. 2006). The superfamily Colubroidea comprises > 80% of the approximately 2900 species of snake currently described (Vidal et al. 2002). Within this taxon, venoms from the Viperinae (vipers) and Crotalinae (pitvipers) subfamilies of Viperidae snakes contain proteins that interfere with the coagulation cascade, the normal hemostatic system, and tissue repair, and human envenomations are often character-

*These authors contributed equally to this work and may both be considered first authors.

Correspondence to: Libia Sanz; *email:* Libia.Sanz@ibv.csic.es

ized by clotting disorders, hypofibrinogenemia, and local tissue necrosis (Markland 1998; Fox and Serrano 2005a). Despite the fact that viperid venoms are complex mixtures of protein components (Fox et al. 2006), venom proteins comprise only a few major protein families, including enzymes (serine proteinases, Zn^{2+}-metalloproteases, L-amino acid oxidase, group II PLA_2) and proteins without enzymatic activity (disintegrins, C-type lectins, natriuretic peptides, myotoxins, CRISP toxins, nerve and vascular endothelium growth factors, cystatin, and Kunitz-type protease inhibitors) (Fry and Wüster 2004; Fry 2005; Markland 1998; Fox et al. 2006; Juárez et al. 2004; Bazaa et al. 2005). Notably, most venom toxins are extensively cross-linked by disulfide bonds and have flourished into functionally diverse, toxin multigene families that exhibit interfamily, intergenus, interspecies, and intraspecific variability. The existence in the same venom of functionally diverse isoforms of the same protein family reflects accelerated Darwinian evolution (Moura da Silva et al. 1996; Ménez 2002; Tani et al. 2002; Ohno et al. 2003). The evolutionary pressure acting to promote high levels of variation in venom proteins may be part of a predator-prey arms race that allows the snake to adapt to a variety of different prey, each of which is most efficiently subdued with a different venom formulation (Daltry et al. 1996). This evolutionary pattern parallels the ''birth-and-death'' model of protein evolution proposed to underpin the evolution of resistance genes in plants (Michelmore and Meyers 1998), the vertebratés adaptative immune responses to protect the organism from a wide range of foreign antigens (Nei et al. 1997; Nei and Rooney 2005), and the evolution of elapid three-finger toxins (Fry et al. 2003). However, details of the molecular events leading to snake venom toxin diversification remain unclear.

Venom Zn^{2+}-dependent metalloproteinases (SVMPs) represent Serpentes-specific hemorrhagic toxins derived from cellular ADAM (A disintegrin and metalloproteinase) proteins (Fry et al. 2006; Moura da Silva et al. 1996; Calvete et al. 2003). Snake venom hemorrhagins have been classified according to their domain structure (Jia et al. 1996; Lu et al. 2005; Fox and Serrano 2005b). The PIII class comprises the closest homologues of cellular ADAMs, which are large multidomain toxins (60–100 kDa) built up by an N-terminal metalloproteinase domain and C-terminal disintegrin-like and cysteine-rich domains. The class PII metalloproteinases (30–60 kDa) contain a disintegrin domain at the carboxyl terminus of the metalloproteinase domain. PI metalloproteinases (20–30 kDa) are single-domain proteins. Disintegrins, a broad family of small (40–100 amino acids), cysteine-rich polypeptides isolated from venoms of vipers and rattlesnakes (Calvete et al. 2005), are released in viper venoms by proteolytic processing of PII SVMP precursors (Shimokawa et al. 1996) or synthesized from short-coding mRNAs (Okuda et al. 2002), and selectively block the function of cell surface adhesive receptors of the integrin family (Calvete et al. 2005; Sanz et al. 2006). Currently, disintegrins can be classified according to their length and number of disulfide bonds (Calvete et al. 2003). The first group includes short disintegrins, composed of 41–51 residues and four disulfide bonds. The second group is formed by the medium-sized disintegrins, which contain about 70 amino acids and six cystine bonds. The third group includes long disintegrins, with an ~84-residue polypeptide cross-linked by seven disulfide bridges. The fourth group is composed of homo- and heterodimers. Dimeric disintegrins contain subunits of about 67 residues with 10 cysteines involved in the formation of four intrachain disulfides and two interchain cystine linkages (Calvete et al. 2000; Bilgrami et al. 2004, 2005). Like many other venom toxins, the integrin inhibitory activity of disintegrins depends on the appropriate pairing of cysteines, which determines the conformation of the inhibitory loop that harbors an active tripeptide located at the apex of a mobile loop protruding 14–17 Å from the protein core (Calvete et al. 2005; Monleón et al. 2003, 2005). The current view is that functional diversification among disintegrins has been achieved during ophidian evolution by amino acid substitutions within the active loop, whereas structural diversification was driven through a disulfide bond engineering mechanism involving the selective loss of pairs of cysteine residues engaged in the formation of disulfide bonds (Calvete et al. 2003). The great sequence and structural diversity exhibited by the different subfamilies strongly suggests that disintegrins, like toxins from other venoms (Duda and Palumbi 1999; Kordis et al. 2002; Ohno et al. 2002), have evolved rapidly by adaptative evolution.

Our earlier studies on venom gland cDNAs encoding disintegrins of *Cerastes vipera*, *Macrovipera lebetina transmediterranea* (Sanz et al. 2006), and *Echis ocellatus* (Juárez et al. 2006a) provided compelling data indicating a common ancestry of the messengers coding for precursors of dimeric disintegrin chains and short disintegrins. In this study we sought to investigate the molecular mechanism underlying the structural diversification of these disintegrins through analysis of the genomic organization of their genes.

Materials and Methods

Extraction of Genomic DNA

Genomic DNA was extracted from fresh tissues of *M. l. transmediterranea* captured in the rocky mountains of northern Tunisia and kept in captivity at the serpentarium of the Institute Pasteur de Tunis (Tunisia) until sacrificed, and from fresh liver of *E. ocellatus* (Kaltungo, Nigeria), of different ages and of both sexes, maintained

Table 1. Forward (F) and reverse (R) primers used for amplification of disintegrin genes from *Macrovipera lebetina transmediterranea* (Ml_X) and *Echis ocellatus* (Eo_X), whose sequences are displayed in Figs. 2 and 3

Primer	Nucleotide sequence	Amplified DNA
F1	5′- ATG AAT TCC GCA AAT CCG TGC-3′	M1_G1, M1_G2
R1	5′-TTA GTC TTT GTA GGG ATT TCT GGG-3′	
F2	5′-CGT GCC ATG GAT TGT ACA ACT GGA CCA TG-3′	M1_G3
R2	5′-G CCT CGA GTA TTA GCC ATT CCC GGG ATA AC-3′	
F3	5′-ATG AAT TCT GCA AAT CCG TGC-3′	Eo_D3
R3	5′-TCA CAT CAA CAC ACT GCC TTT TGC-3′	
F4	5′-GAA CTT TTG CAG AAT TCT G-3′	Eo_C3
R3	5′-TCA CAT CAA CAC ACT GCC TTT TGC-3′	
F5	5′-TGT ACA ACT GGA CCA TGT TGT CG-3′	Eo_RTS
R5	5′-TTA GCC ATT CCC GGG ATA ACT GG-3′	

at the herpetarium of the Liverpool School of Tropical Medicine. The *M. l. transmediterranea* tissues were homogenized in 400 μl of sterile salt buffer (0.4 M NaCl, 110 mM Tris-HCl, pH 8.0, containing 2 mM EDTA) using a Polytron tissue homogenizer for 10–15 s. *Echis ocellatus* liver was ground to a fine powder under liquid nitrogen and the genomic DNA extracted using a Roche DNA isolation kit for cells and tissue containing sodium dodecyl sulfate (SDS; 2% final concentration) and proteinase K (400 μg/ml final concentration). The homogenates were incubated at 55°C overnight. Thereafter, 300 μl of 6 M NaCl (NaCl-saturated H_2O) was added to each sample, and the mixture was vortexed for 30 s at maximum speed and centrifuged for 30 min at 10,000*g*. An equal volume of isopropanol was added to each supernatant, and the sample mixed, incubated at −20°C for 1 h, and centrifuged for 20 min at 4°C and 10,000*g*. The resulting pellets were washed with 70% ethanol, dried, and, finally, resuspended in 300–500 μl sterile distilled H_2O.

DNA Amplification and Sequencing

Disintegrin-encoding DNAs from *M. l. transmediterranea* were amplified by PCR from genomic DNAs using the following pairs of primers, whose sequences are listed in Table 1. For Ml_G1 and Ml_G2, the forward primer was F1, corresponding to the nucleotide sequence coding for the N-terminal amino acid sequence (MNSANPC) of dimeric disintegrin ML-(2,8,15) (SwissProt/TrEMBL accession code AM114016) (Sanz et al. 2006). The reverse primer was R1. For Ml_G3, the forward primer was F2, which contains the sequence coding for the conserved first six residues of the mature short disintegrins jerdostatin (*Protobothrops jerdoni*) (Sanz et al. 2005) (SwissProt/TrEMBL accession code AY262730), CV-short (*Cerastes vipera*) (Q3BK17) (Sanz et al. 2006), and lebestatin (Q3BK14) (*M. l. transmediterranea*) (Sanz et al. 2006) and the CCATGG *Nco*I restriction site. The reverse primer was R2, which includes a STOP codon (TTA), the CTCGAG restriction site for *Xho*I, and the last six C-terminal residues of jerdostatin/CV-short/lebestatin. The Touchdown 60°C/50°C PCR protocol included an initial denaturation step at 95°C for 10 min followed by 4 cycles of denaturation (30 s at 94°C), annealing (30 s at 60°C), and extension (120 s at 72°C); 21 cycles starting with the above conditions and, in subsequent cycles, decreasing the annealing temperature by 0.5°C (reaching 50°C in cycle 21); 10 cycles of denaturation (30 s at 94°C), annealing (30 s at 50°C), and extension (120 s at 72°C); and a final extension for 10 min at 72°C.

Genomic DNA fragments of *E. ocellatus* encoding disintegrins were amplified using primers listed in Table 1. The dimeric disintegrin (Eo_D3) was amplified using the forward primer F3 and the reverse primer R3 complementary to the highly conserved open reading frame (ORF). Amplification of the genomic DNA coding for ocellatusin precursor (Eo_C3) was achieved using the primer F4 designed from the N-terminal of the disintegrin domain and the reverse primer R3. Amplification of the short RTS-containing disintegrin (Eo_RTS) was achieved with the primers F5 (forward) and R5 (reverse). PCR amplification of the *E. ocellatus* genomic DNAs was performed in a 25-μl volume containing 0.34 μg of genomic DNA, dNTP (each at 10 mM), 0.6 μl of 5′ and 3′ primers (10 μM each), 0.25 μl of Ampli Taq Gold polymerase, and buffer provided by the supplier (Roche). PCR amplification was performed using the Touchdown 60°C/50°C PCR protocol as above. The resulting PCR products were directly ligated into a pTOPO vector using the TA cloning kit (Invitrogen) for sequencing.

PCR-amplified genomic DNA fragments from *M. l. transmediterranea* were separated by 1% agarose gel electrophoresis, purified using the Perfect Pre Gel Clean Up kit (Eppendorf, Hamburg, Germany), and cloned in a pGEM-T vector (Promega), which was then used to transform *Escherichia coli* DH5α cells (Novagen, Madison, WI, USA) by electroporation using an Eppendorf 2510 electroporator following the manufacturer's instructions. Positive clones, selected by growing the transformed cells in *Luria* broth (LB) medium containing 10 μg/ml ampicilin, were confirmed by PCR amplification using the above primers and the sequences of the inserts were subjected to sequencing on an Applied Biosystems Model 377 DNA sequencing system.

DNA Sequence Analysis

The nucleotide sequences of Ml_G1, Ml_G2, Ml_G3, Eo_D3, and Eo_C3 were compared to all sequences in the GenBank database (http://www.ncbi.nlm.nih.gov/blast/) (GenBank + EMBL + DDBJ + PDB sequences) using the BLASTN 2.2.13 program (Altschul et al. 1997). Multiple sequence alignment was done with CLUSTALW (Thompson et al. 1994) through the online facility of the Kyoto University Bioinformatics Center (http://clustalw.genome.jp) using default parameters.

Phylogenetic Reconstruction

We have used three methods for phylogeny reconstruction and assessment of the inferred evolutionary relationships based on amino acid sequences. First, we used maximum-likelihood inference as implemented in MEGA 3.1 (Kumar et al. 2001), which optimizes the likelihood function by simultaneously adjusting the topology and branch lengths. We used the JTT (Jones, Taylor, and Thornton 1992) substitution matrix, with a discrete gamma function with eight categories plus invariant sites to account for substitution rate heterogeneity among sites and empirically estimated amino acid frequencies. Second, we employed Bayesian inference with MrBayes 3.1.2. (Ronquist and Huelsenbeck 2003) using the same substitution model (JTT) as in MEGA. The method uses

Markov chain Monte Carlo methods to generate posterior probabilities for each clade represented in the tree. The analysis was performed by running 10^6 generations in four chains, using burn-in at 50,000 generations, and saving every 100th tree. The final tree was rooted using the branch of the human ADAM 7 and 28 molecules as the outgroup.

Accession Numbers

The DNA sequences of clones Ml_G1, Ml_G2, and Ml_G3 from *M. l. transmediterranea* and Eo_D3, Eo_C3, and Eo_RTS from *E. ocellatus* are accessible from the SwissProt/TrEMBL data bank (http://us.expasy.org) under accession codes AM261811, AM261812, AM261813, AM286800, AM286799, and AM286798, respectively.

Results and Discusion

Genomic Organization of Dimeric Disintegrin Subunits

Disintegrin-coding DNAs were PCR-amplified using genomic DNAs as template and a pair of primers complementary for the nucleotide sequences coding for the N- and C-terminal amino acid sequences of dimeric disintegrin subunits from *M. l. transmediterranea* (Mlt) (Sanz et al. 2006) and *E. ocellatus* (Eo) (Juárez et al. 2006a). The 1.2-kb sequences of two such Mlt genes, termed Ml_G1 and Ml_G2 (Fig. 1A), revealed that their ORFs were each interrupted by single introns of 1052 and 1043 bases, respectively (Figs. 2A and 2B). The 1.7-kb sequence of *Echis ocellatus* Eo_D3 (Fig. 1C) coded for a dimeric disintegrin subunit and also contained a single 1006-nucleotide intron (Fig. 3A). The exon-intron-exon structure of the disintegrin domain would support the hypothesis of exon shuffling as a putative mechanism for structural diversification of toxins suggested by Junqueira-de-Azevedo and colleagues (2006) in a recent report on the transcriptome analysis of *Lachesis muta*. The medium-sized disintegrins halystatins 2 and 3 from *Gloydius halys* are produced by alternative splicing of a single gene (GenBank accession code D28871). Addressing this point requires detailed genomic and transcriptomic comparative analyses.

The intragenic regions of Ml_G1, Ml_G2, and Eo_D3 exhibit ~88% sequence identity among themselves and display approximately 90% sequence identity with the full-length (999-base) intron 2 of the gene for prepro-halystatins 2 and 3. The exon-intron organization of prepro-halystatins 2 and 3 and the partial intron sequences (139 nucleotides) of a disintegrin gene from several *Agkistrodon contortrix* subspecies (Soto et al. 2006), which show 92% identity with the 5′ regions of the *M. l. transmediterranea* and the *E. ocellatus* introns (comprising nucleotides 12–155) are the only reported sequences of disintegrin genes (Fig. 4). The partial exon sequences of the *Agkistrodon contortrix* isogenes show the highest similarity with

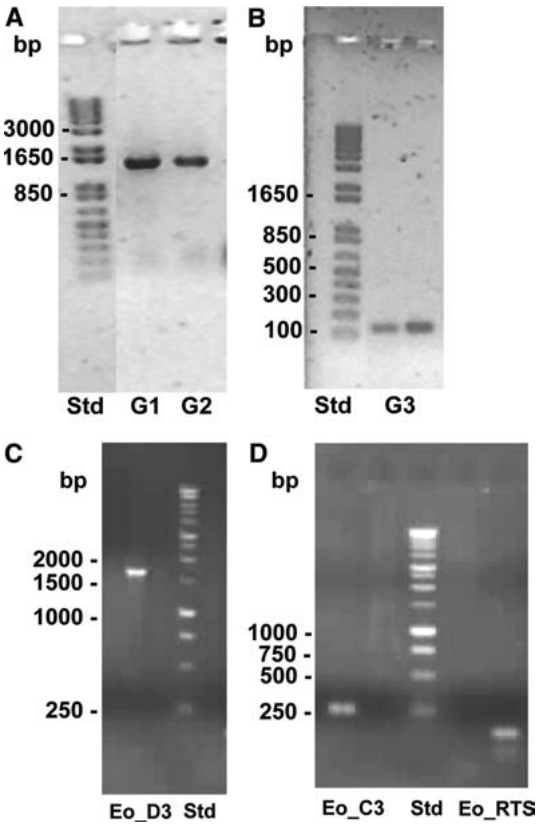


Fig. 1. PCR amplification of disintegrin genes of *M. l. transmediterranea*. One percent agarose gel electrophoretic analysis of the PCR-amplified genomic DNAs from *M. l. transmediterranea* (**A**) Ml_G1 (lane G1) and Ml_G2 (lane G2) and (**B**) Ml_G3 (lane G3). **C**, **D** PCR-amplified genomic DNAs from *E. ocellatus* coding for dimeric disintegrin subunit Eo_D3, short RGD-disintegrin (Eo_C3), and short RTS-disintegrin (Eo_RTS). STD, the "1 kb plus" DNA ladder from Invitrogen was used as a reference for estimating DNA fragment sizes.

cDNA sequences for the α-subunit of acostatin and for the α- and β-subunits of piscivostatin, two dimeric disintegrins from *A. c. contortrix* and *A. piscivorus piscivorus*, respectively (Okuda et al. 2002).

The *M. l. transmediterranea*, *E. ocellatus*, and *G. halys* introns all display the 5′-GTAAG (donor)/3′-AG (acceptor) consensus intron splice site signature (Vicens and Cech 2006) (Figs. 2A and B, 3A, and 4), whereas the introns from the *Agkistrodon contortrix* subspecies genes lack the first 11 nucleotides, which are absolutely conserved in the 5′ end of the *Macrovipera* and the *Gloydius* introns (Fig. 4). Nonetheless, all these introns share equivalent topology within their respective disintegrin genes (Fig. 5). On the other hand, comparison of the exon-intron organization of the gene for the medium-sized disintegrins halystatins 2 and 3 and the genomic DNA coding for the dimeric disintegrin subunits Ml_G1 and Ml_G2 revealed that intron 1 of the former (*G. halys*) disintegrins is absent from the *M. l. transmediterranea* G1 and G2 genes (Fig. 5). Accumulating evidence suggests that the subunits of dimeric disintegrins arose

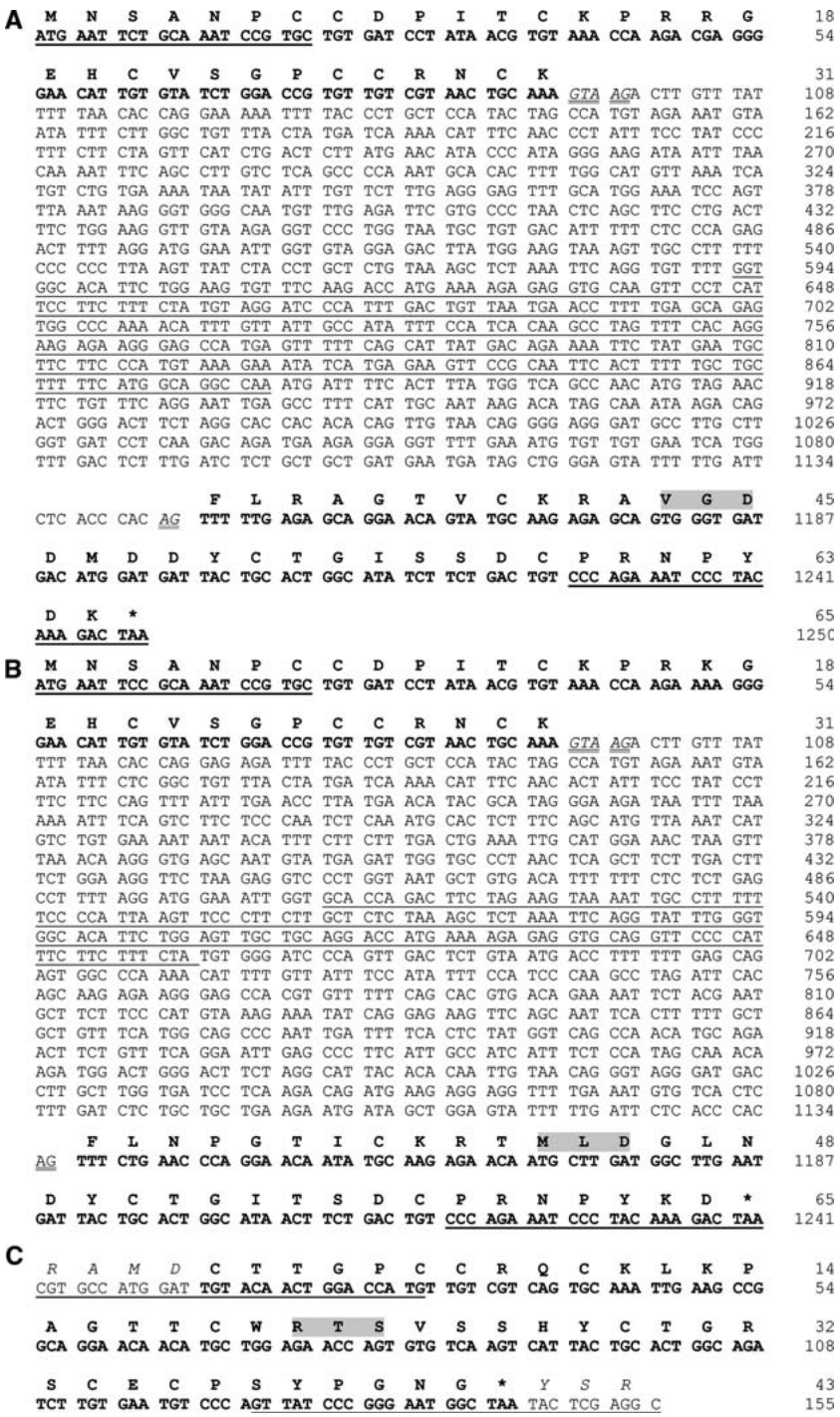


Fig. 2. Disintegrin genes of *M. l. transmediterranea*. Nucleotide sequences of the genomic DNA forMl_G1 (**A**), Ml_G2 (**B**), and Ml_G3 (**C**). The deduced amino acid sequences of exons are shown in the one-letter code. The nucleotide sequences complementary to the primers used for PCR amplification are underlined. In A and B, the 5′-GTAAG (donor)/3′-AG (acceptor) consensus intron splice site signature is in italics and double-underlined, and the intronic sequences overlapping in the 5′ → 3′ and the 3′ → 5′ sequencing directions are underlined. The VGD (A), MLD (B), and RTS (C) integrin binding motifs are depicted on a gray background.

from duplicated medium-sized disintegrin genes (Calvete et al. 2003). Deletions and mutations involving, among others, the codons of the first two cysteine residues, yielded polypeptides with 10 cysteines. Cys-6 and Cys-7, which in monomeric medium-sized disintegrins are disulfide-bonded to the lost cysteines, form interchain disulfide bonds with homologous cysteines from another 10-cysteine-containing disintegrin chain (Calvete et al. 2000; Bilgrami et al. 2004, 2005), giving rise to homo- and heterodimers (Marcinkiewicz et al. 1999a,b; Zhou et al. 2000; Gasmi et al. 2001; Calvete et al. 2002). Beside mutations affecting the exon-coded sequences of disintegrin genes, sequence analysis of dimeric disintegrin subunit genes from different vipers indicates that the loss of intron 1 may represent a conserved feature in the transition from a medium-sized to a subunit of dimeric disintegrin.

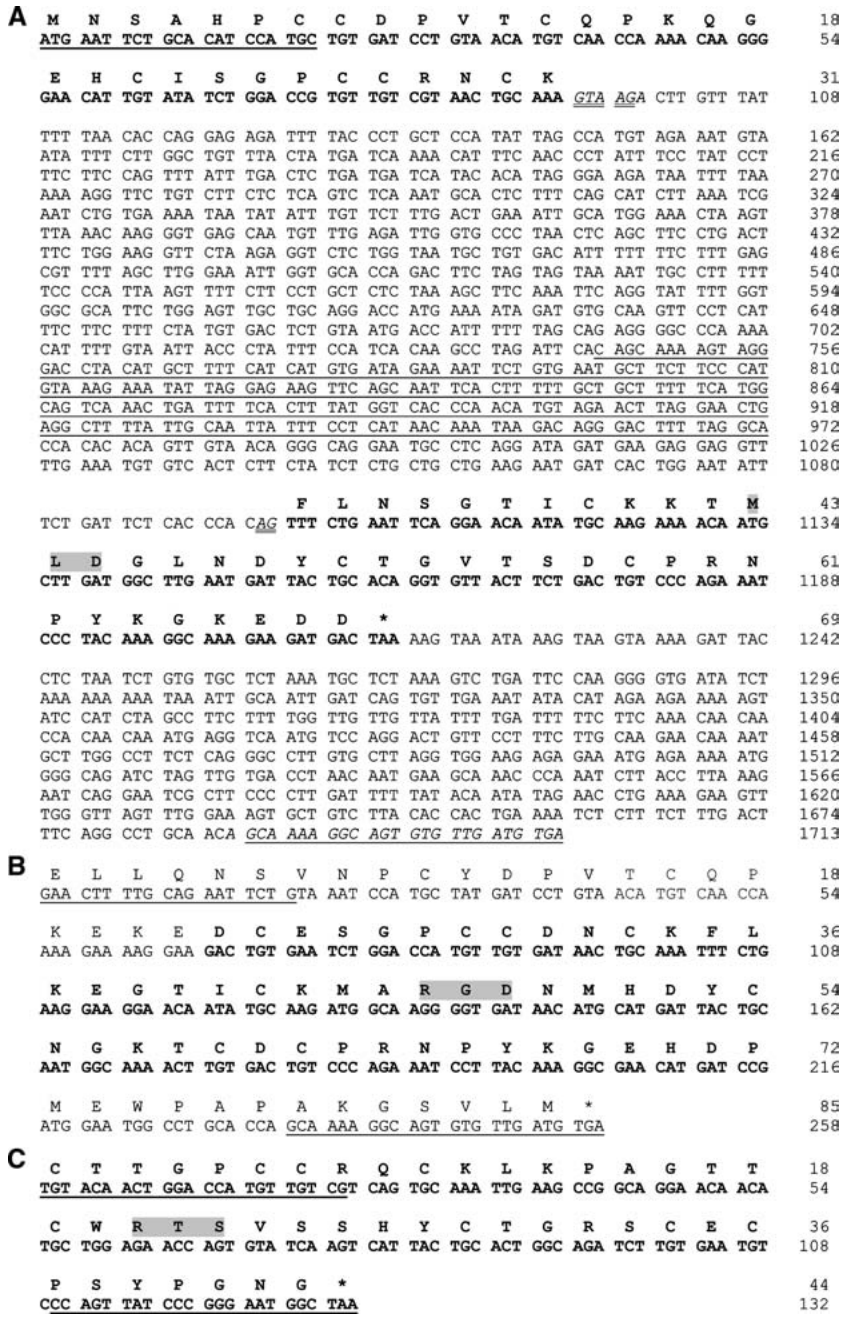


Fig. 3. Disintegrin genes of *E. ocellatus*. Nucleotide sequences of the genomic DNA for Eo_D3 (**A**), Eo_C3 (**B**), and Eo_RTS (**C**). The deduced amino acid sequences of exons are shown in the one-letter code and in boldface. The nucleotide sequences complementary to those of the primers used for PCR amplification are underlined. In A the 5′-GTAAG (donor)/3′-AG (acceptor) consensus intron splice site signature is in italics and double-underlined. Intronic sequences overlapping in the 5′ → 3′ and the 3 → 5′ sequencing directions are underlined. The MLD (A), RGD (B), and RTS (C) integrin binding motifs are depicted on a gray background.

Intronless Genomic Sequences Coding for Short Disintegrins

A 133-bp genomic DNA fragment from *M. l. transmediterranea* (Ml_G3) was amplified (Fig. 1B) using primers complementary to conserved regions of the short disintegrins lebestatin (*M. l. transmediterranea*), CV-short (*Cerastes vipera*) (Sanz et al. 2006), and jerdostatin (*Protobothrops jerdoni*) (Sanz et al. 2005). This DNA fragment corresponded to a small intronless genomic sequence coding for a full-length RTS-disintegrin (Fig. 2C) identical to mature CV-short and jerdostatin. Ml-G3 and lebestatin show 89% amino acid sequence identity. An identical RTS-disintegrin sequence (Eo_RTS) was amplified from genomic DNA of *E. ocellatus* using a different set of primers (Figs. 1D and 3C). This striking finding was consistently confirmed in more than 20 DNA sequence experiments from independent clones. Thus, using the same set of primers, we (Sanz et al., in preparation) have amplified intronless jerdostatin-like genes from genomic DNAs of a number of species classified into very diverse genera from the subfamilies Viperinae (pitless vipers: *Daboia russelli*, *Bitis arietans*) and Crotalinae (pit vipers: *P. jerdoni*, *G. halys*, *Sistrurus catenatus*

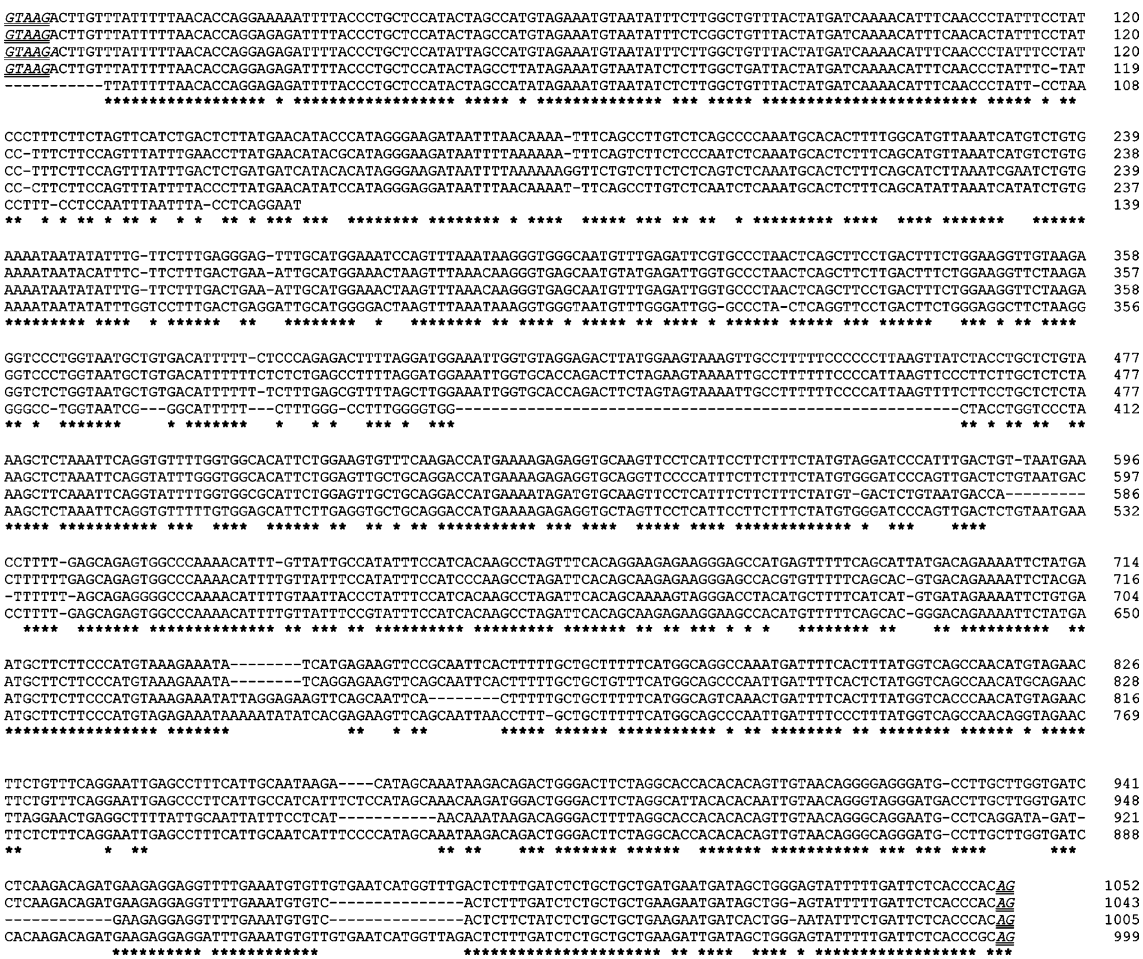


Fig. 4. Comparison of intronic sequences. Alignment of the nucleotide sequences of the introns of Ml_G1, Ml_G2, and Eo_D3 with that of intron 2 of prepro-halystatins 2 and 3 from the Chinese water mocassin (*G. halys pallas*) (GenBank accession code D28871) and the partial intronic sequence of a disintegrin gene from several *Agkistrodon contortrix* subspecies (Soto et al. 2006). Identical nucleotides are labeled with an asterisk below the multiple alignment. The 5′-GTAAG (donor)/3′-AG (acceptor) consensus intron splice site signature is in italics and double-underlined.

catenatus, *Crotalus viridis*, *P. mucrosquamatus*) (http://www.embl-heidelberg.de/~uetz/families/Viperidae.html).

The short RGD-disintegrin ocellatusin (*E. ocellatus*) is also transcribed from an intronless genomic 258-bp DNA sequence (Eo_C3) (Figs. 1D and 3B). The nucleotide sequence of Eo_C3 is identical to the region comprising nucleotides 1228–1485 of the long precursor (PII metalloprotease) of ocellatusin (clone Eo_00006; Juárez et al. 2006a). Analysis of cDNAs from *M. l. transmediterranea* and *E. ocellatus* venom gland libraries encoding disintegrins argued strongly for a common ancestry of the messengers of short disintegrins and those for precursors of dimeric disintegrin chains (Juárez et al. 2006a; Sanz et al. 2006). In line with this evidence, our current findings indicate that the evolutionary pathway leading to the emergence of short disintegrins involved the removal of all intronic sequences (Fig. 6).

Loss of Introns Along the Diversification Pathway of Disintegrins

A survey of complete genome sequences (http://www.ensembl.org;http://www.ncbi.nlm.nih.gov/BLAST) revealed that the genes for ADAM proteins from a number of species, including zebra fish (*Danio rerio*), mouse (*Mus musculus*), rat (*Rattus norvegicus*), chicken (*Gallus gallus*), and human (*Homo sapiens*), contain three topologically conserved introns within their disintegrin-like domains (Fig. 5). ADAMs represent the closest homologues of PIII snake venom proteins, which in turn are likely to represent the ancestors of the PII disintegrin family (Moura da Silva et al. 1996; Calvete et al. 20032005; Calvete 2005). Noteworthily, all known ADAM and disintegrin introns share the 5′-GTAAG (donor)/3′-AG (acceptor) consensus splice site signature (Figs. 5 and 6). Moreover, the insertion sites of introns 1 and 2 are conserved in the genes for

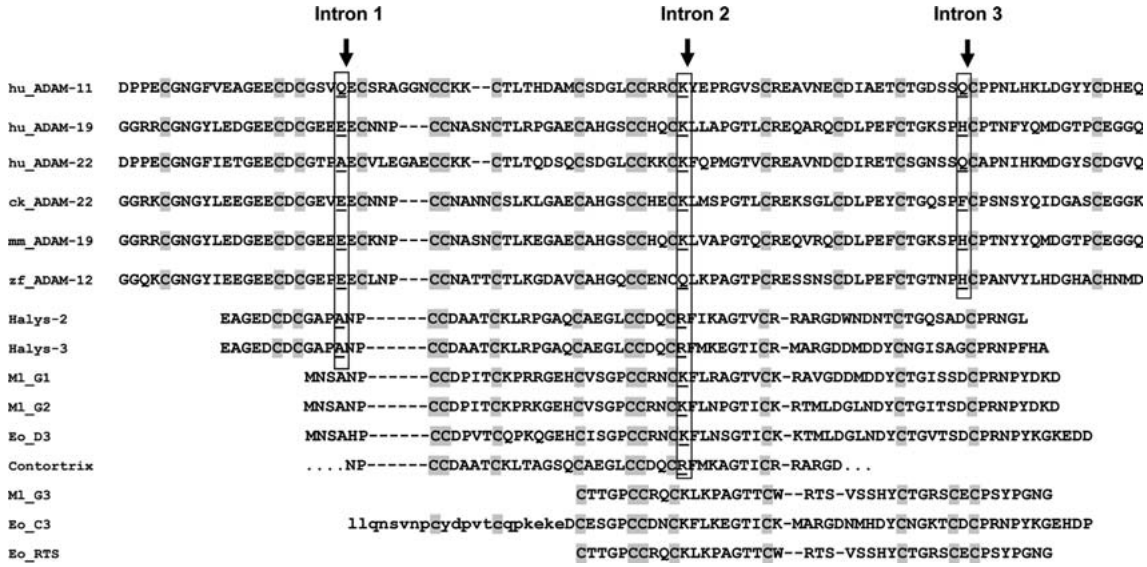


Fig. 5. Comparison of amino acid sequences and organization of disintegrin genes. Alignment of the genomic DNA-deduced amino acid sequences of dimeric disintegrin subunits Ml_G1, Ml_G2, and Eo_D3 and the short disintegrins Ml_G3 and Eo_C3 from *M. l. transmediterranea* ("Ml_X") and *E. ocellatus* ("Eo_X") with representative disintegrin-like domains from ADAM molecules of other vertebrate taxa (hu, human; ck, chicken; mm, mouse; zf, zebra fish) and snake venom disintegrins whose gene organization has been reported: Halys-2 and Halys-3, disintegrin domains of halystatins 2 and 3 from the Chinese water mocassin (*Gloydius halys pallas*; GenBank accession code D28871); Contortrix, partial sequence of a disintegrin gene from several *Agkistrodon contortrix* subspecies (Soto et al. 2006). The amino acid residues preceding the conserved insertion position of introns 1, 2, and 3 (marked with arrows) in the respective disintegrin/disintegrin-like genes are underlined and boxed. Cysteine residues are depicted on a gray background. Accession codes: hu_ADAM-11 (NC_000017.9), hu_ADAM-19 (NC_000005.8), hu_ADAM-22 (NC_000007.12), ck_ADAM-22 (http://www.ensembl.org; ENSGALP00000014587), mm_ADAM-19 (AK147217), zf_ADAM-22 (NC_007128.1).

ADAMs and for the medium-sized snake disintegrins halystatins 2 and 3, and in addition, the insertion position of intron 2 is also conserved in the genes for dimeric disintegrin subunits Ml_G1, Ml_G2, and Eo_D3 (Figs. 5 and 6). On the other hand, intron 3 has been removed from the halystatin gene; introns 1 and 3 are not present in the genes for Ml_G1, Ml_G2, and Eo_D3; and the short disintegrins Ml_G3 and Eo_C3 (ocellatusin) are encoded by intronless genes (Figs. 5 and 6).

Phylogenetic relationships among disintegrins inferred using maximum likelihood and Bayesian analysis (Fig. 7) indicate that the disintegrin-like domains of PIII snake venom metalloproteinases and those of cellular ADAMs are the closest homologues, and are in line with the view that PIII SVMPs diverged from an ADAM precursor (Moura da Silva et al. 1996). PIII SVMPs are modular proteins containing N-terminal Zn^{2+}-metalloproteinase, disintegrin-like, and C-terminal cysteine-rich domains. Gene duplication and removal of the region encoding the cysteine-rich domain generated the PII SVMP genes (Moura da Silva et al. 1996; Juárez et al. 2006b). The phylogenetic tree inferred for the disintegrin family (Fig. 7) also supports our hypothesis that the structural diversity of PII disintegrins has been achieved during evolution through mutations causing the loss of pairs of cysteine residues engaged in the formation of disulfide bonds, generating successively the precursors of long, medium-sized, dimeric, and short disintegrins as depicted schematically in Fig. 6A (Calvete et al. 2003; Juárez et al. 2006a). Our current results showing the sequential loss of introns along the diversification pathway of disintegrins provide additional evidence in favor of our hypothesis that a minimization of both the gene organization and the protein structure underpins the evolution of the snake venom disintegrin family (Fig. 6). Challenges for future investigations are to identify the molecular machinery responsible for, and to dissect the individual steps along, the transformation pathway of disintegrins from ADAM PIII disintegrin-like domains to snake venom short PII disintegrins.

Acknowledgments. This study was partly financed by Grant BFU2004-01432/BMC from the Ministerio de Educación y Ciencia, Madrid, Spain, and by Spanish-Tunisian Cooperation Programme 29p/02 of the Agencia Española de Cooperación Internacional (AECI). P.J. is the recipient of a predoctoral fellowship from the Formación de Personal Investigador (FPI) program from the Ministerio de Educación y Ciencia, Madrid. The authors thank Iñaki Comas (Instituto Cavanilles de Biodiversidad y Biología Evolutiva, Universidad de Valencia) for helping with phylogenetic analysis.

References

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

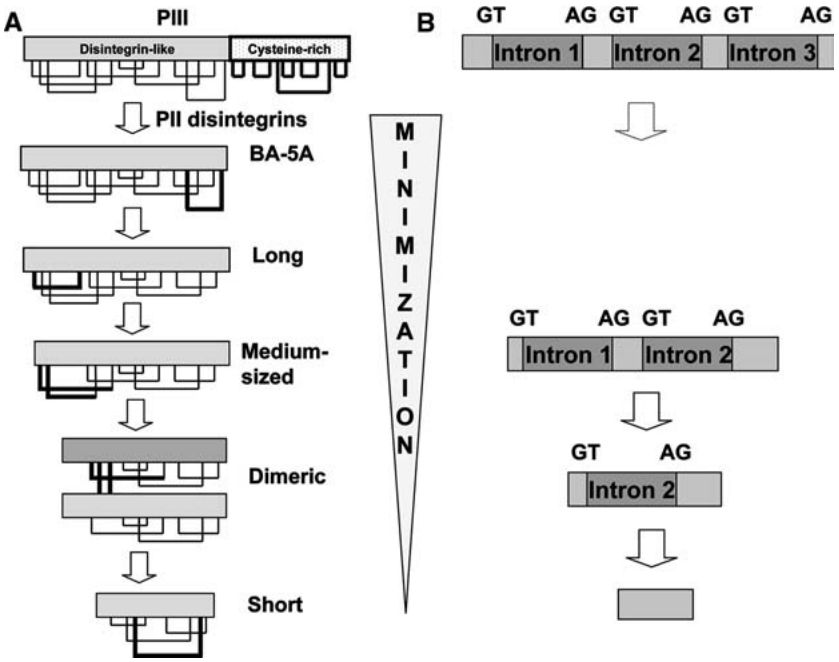


Fig. 6. Minimization of protein structure and gene organization. **A** Scheme of the domain organization, disulfide bond patterns, and proposed evolutionary pathway from the PIII disintegrin/cysteine-rich proteins to short disintegrins (Calvete et al. 2003, 2005; Calvete 2005; Juárez et al. 2006a, b). Structural features (the cysteine-rich domain of PIII disintegrin-like molecules, and class-specific disulfides) lost along the disintegrin diversification pathway are highlighted with thick lines. **B** Scheme of the conserved exon-intron organization of known disintegrin-like domains of vertebrate ADAM proteins and of the medium-sized disintegrins halystatins 2 and 3, the dimeric disintegrin subunits Ml_G1, Ml_G2, and Eo_D3, and the short-disintegrins Ml_G3 and Eo_C3. *GT...AG*, denotes the 5-***GT***AAG (donor)/3-***AG*** (acceptor) consensus intron splice site signature conserved in all known disintegrin-like and disintegrin genomic DNAs. The concept of minimization of both the protein and the gene structures along the diversification pathway of disintegrins is highlighted.

Bazaa A, Marrakchi N, El Ayeb M, Sanz L, Calvete JJ (2005) Snake venomics: comparative analysis of the venom proteomes of the Tunisian snakes *Cerastes cerastes*, *Cerastes vipera* and *Macrovipera lebetina*. Proteomics 5:4223–4235

Bilgrami S, Tomar S, Yadav S, Kaur P, Kumar J, Jabeen T, Sharma S, Singh TP (2004) Crystal structure of schistatin, a disintegrin homodimer from saw-scaled viper (*Echis carinatus*) at 25 Å resolution. J Mol Biol 341:829–837

Bilgrami S, Yadav S, Sharma S, Perbandt M, Betzel C, Singh TP (2005) Crystal structure of the disintegrin heterodimer from saw-scaled viper (*Echis carinatus*) at 19 Å resolution. Biochemistry 44:11058–11066

Calvete JJ (2005) Structure-function correlations of snake venom disintegrins. Curr Pharm Des 11:829–835

Calvete JJ, Jürgens M, Marcinkiewicz C, Romero A, Schrader M, Niewiarowski S (2000) Disulfide bond pattern and molecular modelling of the dimeric disintegrin EMF-10, a potent and selective integrin $\alpha_5\beta_1$ antagonist from *Eristocophis macmahoni* venom. Biochem J 345:573–581

Calvete JJ, Fox JW, Agelan A, Niewiarowski S, Marcinkiewicz C (2002) The presence of the WGD motif in CC8 heterodimeric disintegrin increases its inhibitory effect on $\alpha_{IIb}\beta_3$, $\alpha_v\beta_3$, and $\alpha_5\beta_1$ integrins. Biochemistry 41:2014–2021

Calvete JJ, Moreno-Murciano MP, Theakston RDG, Kisiel DG, Marcinkiewicz C (2003) Snake venom disintegrins: novel dimeric disintegrins and structural diversification by disulphide bond engineering. Biochem J 372:725–734

Calvete JJ, Marcinkiewicz C, Monleón D, Esteve V, Celda B, Juárez P, Sanz L (2005) Snake venom disintegrins: evolution of structure and function. Toxicon 45:1063–1074

Daltry JC, Wüster W, Thorpe RS (1996) Diet and snake venom evolution. Nature 379:537–540

Duda TF Jr, Palumbi SR (1999) Developmental shifts and species selection in gastropods. Proc Natl Acad Sci USA 96:6820–6823

Fox JW, Serrano SMT (eds) (2005a) Special Issue: Snake Toxins and Hemostasis. Toxicon 45:951–1181

Fox JW, Serrano SMT (2005b) Structural considerations of the snake venom metalloproteinases, key members of the M12 reprolysin family of metalloproteinases. Toxicon 45:969–985

Fox JW, Ma L, Nelson K, Sherman NE, Serrano SMT (2006) Comparison of indirect and direct approaches using ion-trap and Fourier transform ion cyclotron resonance mass spectrometry for exploring viperid venom proteomes. Toxicon 47:700–714

Fry BG (2005) From genome to ''venome'': molecular origin and evolution of the snake venom proteome inferred from phylogenetic analysis of toxin sequences and related body proteins. Genome Res 15:403–420

Fry BG, Wüster W (2004) Assembling an arsenal: origin and evolution of the snake venom proteome inferred from phylogenetic analysis of toxin sequences. Mol Biol Evol 21:870–883

Fry BG, Wüster W, Kini RM, Brusic V, Khan A, Venkataraman D, Rooney AP (2003) Molecular evolution and phylogeny of elapid snake venom three-finger toxins. J Mol Evol 57:110–129

Fry BG, Vidal N, Norman JA, Vonk FJ, Scheib H, Ramjan SF, Kuruppu S, Fung K, Hedges SB, Richardson MK, Hodgson WC, Ignjatovic V, Summerhayes R, Kochva E (2006) Early evolution of the venom system in lizards and snakes. Nature 439:584–588

Gasmi A, Srairi N, Guermazi S, Dkhil H, Karoui H, El Ayeb M (2001) Amino acid structure and characterization of a heterodimeric disintegrin from *Vipera lebetina* venom. Biochim Biophys Acta 1547:51–56

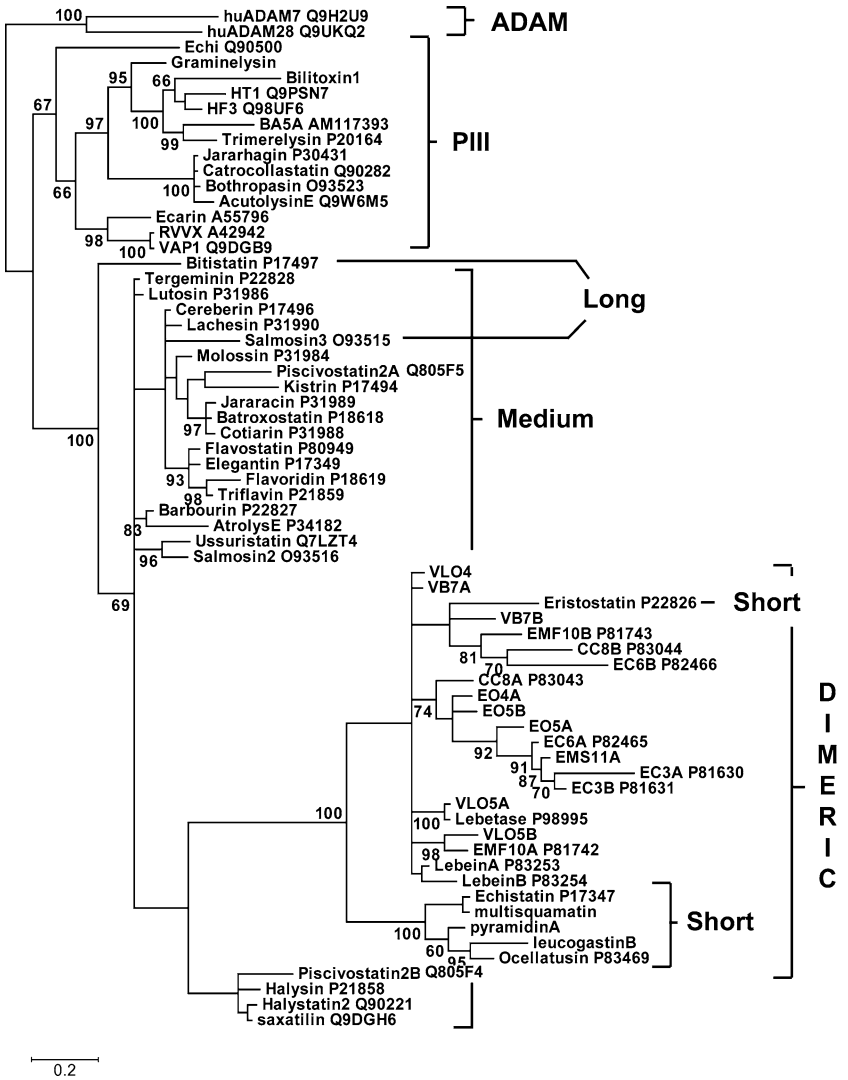


Fig. 7. Phylogenetic tree for the disintegrin family. Evolutionary relationships among disintegrins were inferred using maximum-likelihood and Bayesian methods. The analyses were based on the amino acid sequences of PIII disintegrin-like and PII long, medium-sized, dimeric, and short disintegrin domains displayed in Fig. 3 of Calvete et al. (2003). When available, database accession numbers are indicated. The sequences of the short disintegrins multisquamatin, pyramidinA, and leucogastin are from Okuda et al. (2001). The primary structures of the dimeric disintegrin subunits EMS11A, VLO4, VLO5A, VLO5B, VA6, VB7A, and VB7B are reported by Calvete et al. (2003). Bilitoxin-1 is from Nikai et al. (2000), and the sequence of the disintegrin-like domain of graminelysin is from Wu et al. (2001). For rooting the tree, the branch of disintegrin-like domains of the human ADAM-7 and ADAM-28 molecules was used as the outgroup. Nodes with confidence values $>60\%$ are indicated.

Jia L-G, Shimokawa K-i, Bjarnason JB, Fox JW (1996) Snake venom metalloproteinases: structure, function and relationship to the ADAMs family of proteins. Toxicon 34:1269–1276

Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci 8:275–282

Juárez P, Sanz L, Calvete JJ (2004) Snake venomics: characterization of protein families in *Sistrurus barbouri* venom by cysteine mapping, N-terminal sequencing, and tandem mass spectrometry analysis. Proteomics 4:327–338

Juárez P, Wagstaff SC, Sanz L, Harrison RA, Calvete JJ (2006a) Molecular cloning of *Echis ocellatus* disintegrins reveals non-venom-secreted proteins and a pathway for the evolution of ocellatusin. J Mol Evol 63:183–193

Juárez P, Wagstaff SC, Oliver J, Sanz L, Harrison RA, Calvete JJ (2006b) Molecular cloning of disintegrin-like transcript BA-5A from a *Bitis arietans* venom gland cDNA library: a putative intermediate in the evolution of the long chain disintegrin bitistatin. J Mol Evol 63:142–152

Junqueira-de-Azevedo ILM, Ching ATC, Carvalho E, Faria F, Nishiyama NY, Ho PL, Diniz MRV (2006) *Lachesis muta* (Viperidae) cDNAs reveal diverging pit viper molecules and scaffolds typical of cobra (Elapidae) venoms: implications for snake toxin repertoire evolution. Genetics 173:877–889

Kordis D, Krizaj I, Gubensek F (2002) Functional diversification of animal toxins by adaptative evolution. In: Ménez A (ed) Perspectives in molecular toxinology. John Wiley & Sons, Chichester, UK, pp 401–419

Kumar S, Tamura K, Jacobsen IB, Nei M (2001) Mega2: Molecular Evolutionary Genetic Analysis software. Bioinformatics 17:1244–1245

Lu X, Lu D, Scully MF, Kakkar VV (2005) Snake venom metalloproteinase containing a disintegrin-like domain, its structure-

activity relationships at interacting with integrins. Curr Med Chem Cardiovasc Hematol Agents 3:249–260

Marcinkiewicz C, Calvete JJ, Marcinkiewicz MM, Raida M, Vijay-Kumar S, Huang Z, Lobb RR, Niewiarowski S (1999a) EC3, a novel heterodimeric disintegrin from *Echis carinatus* venom, inhibits α4 and α5 integrins in an RGD-independent manner. J Biol Chem 274:12468–12473

Marcinkiewicz C, Calvete JJ, Vijay-Kumar S, Marcinkiewicz MM, Raida M, Schick P, Lobb RR, Niewiarowski S (1999b) Structural and functional characterization of EMF10, a heterodimeric disintegrin from *Eristocophis macmahoni* venom that selectively inhibits α5β1 integrin. Biochemistry 38:13302–13309

Markland FS (1998) Snake venoms and the hemostatic system. Toxicon 36:1749–1800

Ménez A (2002) Perspectives in molecular toxinology. John Wiley & Sons, Chichester, UK

Michelmore RW, Meyers BC (1998) Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. Genome Res 8:1113–1130

Monleón D, Moreno-Murciano MP, Kovacs H, Marcinkiewicz C, Calvete JJ, Celda B (2003) Concerted motions of the integrin-binding loop and the C-terminal tail of the non-RGD disintegrin obtustatin. J Biol Chem 278:45570–45576

Monleón D, Esteve V, Kovacs H, Calvete JJ, Celda B (2005) Conformation and concerted dynamics of the integrin-binding site and the C-terminal region of echistatin revealed by homonuclear NMR. Biochem J 387:57–66

Moura-Da-Silva AM, Theakston RDG, Crampton JM (1996) Evolution of disintegrin cysteine-rich and mammalian matrix-degrading metalloproteinases: gene duplication and divergence of a common ancestor rather than convergent evolution. J Mol Evol 43:263–269

Nei M, Rooney AP (2005) Concerted and birth-and-death evolution of multigene families. Annu Rev Genet 39:121–152

Nei M, Gu X, Sitnikova T (1997) Evolution by the birth-and-death process in multigene families of the vertebrate immune system. Proc Natl Acad Sci USA 94:7799–7806

Nikai T, Taniguchi K, Komori Y, Masuda K, Fox JW, Sugihara H (2000) Primary structure and functional characterization of bilitoxin-1, a novel dimeric P–II snake venom metalloproteinase from *Agkistrodon bilineatus* venom. Arch Biochem Biophys 378:6–15

Ohno M, Ogawa T, Oda–Ueda N, Chijiwa T, Hattori S (2002) Accelerated and regional evolution of snake venom gland isozymes. In: Ménez A (ed) Perspectives in molecular toxinology. John Wiley & Sons, Chichester, UK, pp 387–401

Ohno M, Chijiwa T, Oda-Ueda N, Ogawa T, Hattori S (2003) Molecular evolution of myotoxin phospholipases A_2 from snake venom. Toxicon 42:841–854

Okuda D, Nozaki C, Sekiya F, Morita T (2001) Comparative biochemistry of disintegrins isolated from snake venom: consideration of the taxonomy and geographical distribution of snakes in the genus *Echis*. J Biochem 129:615–620

Okuda D, Koike H, Morita T (2002) A new gene structure of the disintegrin family: a subunit of dimeric disintegrin has a short coding region. Biochemistry 41:14248–14254

Ronquist F, Huelsenbeck JP (2003) MrBayes3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574

Sanz L, Chen RQ, Pérez A, Hilario R, Juárez P, Marcinkiewicz C, Monleón D, Celda B, Xiong YL, Pérez-Payá E, Calvete JJ (2005) cDNA cloning and functional expression of jerdostatin, a novel RTS-disintegrin from *Protobothrops jerdoni* and a specific antagonist of the $\alpha_1\beta_1$ integrin. J Biol Chem 280:40714–40722

Sanz L, Bazaa A, Marrakchi N, Pérez A, Chenik M, Bel Lasfer Z, El Ayeb M, Calvete JJ (2006) Molecular cloning of disintegrins from *Cerastes vipera* and *Macrovipera lebetina transmediterranea* venom gland cDNA libraries: insight into the evolution of the snake venom integrin-inhibition system. Biochem J 395:385–392

Shimokawa K, Jai L-G, Wang X-M, Fox JW (1996) Expression, activation and sequencing of the recombinant snake venom metalloproteinase, pro-atrolysin E. Arch Biochem Biophys 335:283–294

Soto JG, Powell RL, Reyes SR, Wolana L, Swanson LJ, Sanchez EE, Perez JC (2006) Genetic variation of a disintegrin gene found in the American copperhead snake (*Agkistrodon contortrix*). Gene 373:1–7

Tani A, Ogawa T, Nose T, Nikandrov NN, Deshimaru M, Chijiwa T, Chang CC, Fukumaki Y, Ohno M (2002) Characterization, primary structure and molecular evolution of anticoagulant protein from *Agkistrodon actus* venom. Toxicon 40:803–813

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Vicens Q, Cech TR (2006) Atomic level architecture of group I introns revealed. Trends Biochem Sci 31:41–51

Vidal N (2002) Colubroid systematics: evidence for an early appearance of the venom apparatus followed by extensive evolutionary tinkering. J Toxicol Toxin Rev 21:21–41

Wu WB, Chang SC, Liau MY, Huang TF (2001) Purification, molecular cloning and mechanism of action of graminelysin I, a snake-venom-derived metalloproteinase that induces apoptosis of human endothelial cells. Biochem J 357:719–728

Zhou Q, Hu P, Ritter MR, Swenson SD, Argounova S, Epstein AL, Markland FS (2000) Molecular cloning and functional expression of contortrostatin, a homodimeric disintegrin from southern copperhead snake venom. Arch Biochem Biophys 375:278–288

4. DISCUSIÓN

Los venenos de las serpientes actuales se originaron en glándulas salivares modificadas (glándula de Duvernoy) compuestas de túbulos muy ramificados alojados en una masa de tejido conectivo en la mandíbula superior. La composición y acción biológica de estos venenos han sido refinadas a lo largo de varias decenas de millones de años de evolución. La existencia de familias multigénicas de toxinas está ampliamente documentada en numerosos estudios bioquímicos de venenos de diversas serpientes. Se acepta generalmente que ello es el resultado de duplicaciones génicas y posterior evolución acelerada. El concepto de evolución acelerada fue acuñado por Motonori Ohno y colegas (Ohno et al. 2002) analizando los genes de isoenzimas de PLA_2s de diversas especies de serpientes del género *Trimeresurus* que habitan en diferentes islas del archipielago japonés. Si definimos K_N como el número de sustituciones de nucleótidos por sitio en regiones intrónicas, y K_S y K_A como los números de sustituciones de nucleótidos por sitio sinónimo y no sinónimo en regiones codificantes (exones), respectivamente, la evolución acelerada en el seno de la familia multigénica de las PLA_2s se caracteriza por: a) K_N ~ 1/4 K_S, y b) K_A/K_S ~/> 1. La condición (a) indica que, al contrario de lo que se observa en genes ordinarios, los intrones de los genes que codifican $PLA2_s$ de venenos de *Trimeresurus* varían menos que los exones. También el cociente K_A/K_S es mucho menor en genes ordinarios (~0.2) que en los genes para PLA_2s (>0.75). La elevada tasa de mutaciones que conllevan sustituciones de aminoácidos en regiones codificantes indica que el gen duplicado, al no estar sometido a una presión de selección que preserve la función ancestral (garantizada por una copia del gen), puede adaptarse a una nueva función (adaptación Darwiniana positiva). El fenómeno de evolución acelerada ha sido también descrito para genes de diversas toxinas, tanto de serpientes (serinproteasas (Deshimaru et al. 1996); toxinas 3F ("three-finger") (Fry et al. 2003; Ohno et al. 1998); SVMPs (Kordis et al. 2002); inhibidores de proteasas tipo Kunitz y BPTI (Kordis et al. 2002; Zupunski et al. 2003) como de moluscos del género *Conus* (conotoxinas) (Olivera et al. 1999) y de escorpiones del género *Tityus* (Becerril

et al. 1997). En el caso de la familia de las disintegrinas, sólo recientemente se han podido analizar secuencias génicas de disintegrinas del género *Crotalus* (Soto et al. 2006a; Soto et al. 2006b). Los valores de K_A/K_S obtenidos para las regiones exónicas de todos los pares de genes homólogos analizados fueron en la mayoría de los casos entre 0.9 y 1.2, con valores máximos para los pares Atroxatina vs Mojastina (3.5) y Atroxatina vs Atrolisina E (2.4). Analizando las secuencias de nucleótidos de las disintegrinas cortas y diméricas de *Echis ocellatus* y *Macrovipera lebetina transmediterranea* descritas en el artículo 4 obtuvimos los valores mostrados en la Tabla 1.

Pares de genes	K_S	K_A	K_A/K_S
Eo_C3 vs Eo_D3	0,15	0,26	1,69
Eo_C3 vs Eo_RTS	0,18	0,48	2,59
Eo_C3 vs Ml_G1	0,20	0,21	1,05
Eo_C3 vs Ml_G2	0,17	0,23	1,32
Eo_C3 vs Ml_G3	0,18	0,44	2,59
Eo_D3 vs Eo_RTS	0,29	0,44	1,50
Eo_D3 vs Ml_G1	0,14	0,15	1,03
Eo_D3 vs Ml_G2	0,18	0,07	0,38
Eo_D3 vs Ml_G3	0,29	0,44	1,50
Eo_RTS vs Ml_G1	0,23	0,44	1,90
Eo_RTS vs Ml_G2	0,17	0,43	2,48
Ml_G1 vs Ml_G2	0,06	0,10	1,61
Ml_G1 vs Ml_G3	0,23	0,44	1,90
Ml_G2 vs Ml_G3	0,17	0,43	2,48

Tabla 1. Análisis de la relación K_A/K_s para pares de genes de disintegrinas de *Echis ocellatus* y *Macrovipera lebetina transmediterranea*. Los parámetros de variación génica fueron calculados con el programa Mega 3.1.

Todos los pares de genes comparados tienen una relación K_A/K_S mayor que 1, excepto Eo_D3 vs Ml_G2, dos disintegrinas cuya similitud de secuencia nucleotídica es del 90.1%. Nuestros resultados corroboran y amplían la hipótesis de que, al igual que se ha descrito para otras toxinas, la diversificación estructural (y funcional) en el seno de la familia de las disintegrinas ha surgido mediante evolución acelerada.

La evolución adaptativa de toxinas de familias multigénicas amplía las posibilidades de supervivencia de un organismo sometido a nichos ecológicos cambiantes en cuanto a clima o tipo de presa predominante. Sin embargo, los mecanismos moleculares básicos del reclutamiento, transformación de proteínas ordinarias en toxinas y de la evolución acelerada de toxinas de familias multigénicas son esencialmente desconocidos.

Además de la trascendencia en el campo de la Evolución Molecular, el conocimiento de los mecanismos de diversificación estructural y funcional de proteínas tiene un indudable interés biotecnológico. El establecimiento de correlaciones estructura-función es básico para el diseño racional de drogas de utilidad clínica, como también para el desarrollo de compuestos con actividad anti-toxina para casos de envenenamiento. Así, en el caso de venenos de serpientes, además de los anteriormente citados antiagregantes de plaquetas Tirofiban (Aggrastat®) y Eptifibatide (Integrilin®) basados en la secuencia RGD, Alfimeprase® (Nuvelo-R&D), una SVMP fibrinolítica recombinante derivada de fibrolasa del veneno de *Agkistrodon contortrix contortrix* (Jones et al. 2001), ha completado la fase II de ensayo clínico, y Captopril Cinfa® (Bristol-Myers), el primer inhibidor de la enzima convertidora de la angiotensina I que produce una relajación de los vasos sanguíneos y reduce la presión arterial, se basó en la estructura de los péptidos potenciadores de la bradiquinina. El aprovechamiento de los secretos ocultos en los venenos de serpientes pasa por una caracterización exhaustiva de sus componentes.

4.1.VENÓMICA: CARACTERIZACIÓN PROTEÓMICA DE VENENOS DE SERPIENTES

Cuando decidimos iniciar la caracterización de la composición proteica de venenos de serpientes no existía en la literatura ningún estudio similar. Había numerosos trabajos que describían la caracterización bioquímica y funcional de toxinas aisladas de diferentes venenos de serpientes de las familias Viperidae y Elapidae, pero un áspecto tan básico como la concentración relativa de estas toxinas en el veneno permanecía por lo general indeterminada. La composición proteica (proteoma) y la abundancia relativa de las diferentes familias de toxinas en un veneno particular es relevante tanto para entender los efectos biológicos de dicho veneno, como para -por ejemplo- delimitar el mínimo número de epítopos cuyo bloqueo sería teoricamente suficiente para neutralizar la toxicidad del veneno. Los antivenenos actuales se basan en la inmunización de caballos, camellos o llamas con dosis subletales de veneno completo. El antisuero resultante contiene anticuerpos frente a epítopos inmunodominantes que no tienen por qué representar epítopos neutralizantes. Tomemos como ejemplo una familia multigénica de SVMP PIII. El sistema inmunitario del animal generará un gran número de anticuerpos frente a una proteína multidominio de más de 500 aminoácidos, muchos de los cuales probablemente carecerán de actividad neutralizante. En teoría, sin embargo, bastaría con producir unos pocos anticuerpos frente a epítopos funcionales estructuralmente conservados en todas las isoenzimas de la familia para neutralizar la actividad catalítica de todas las metaloproteasas de la familia. En este sentido, el grupo que dirige el Dr. Rob Harrison en la Escuela de Medicina Tropical de la Universidad de Liverpool (con quién hemos colaborado en la realización del trabajo presentado en los Artículos 2, 3 y 4) está evaluando un prometedor método de inmunoprotección frente a venenos de las serpientes africanas de mayor importancia médica, *Echis ocellatus*

(Nigeria), *Cerastes cerastes* (Egipto), y *Bitis arietans* (Ghana), basado en transfectar, utilizando una pistola de genes ("GeneGun"), células epidérmicas con un gen sintético formado por una serie lineal de epítopos neutralizantes de las toxinas mayoritarias dispuestos sobre partículas de oro de ~ 1.6 μm de diámetro (Harrison 2004). Actualmente la selección de los epítopos se realiza mediante técnicas bioinformáticas utilizando como base de datos el conjunto de secuencias obtenidas por aproximación transcriptómica (Wagstaff et al. 2006). Utilizando un gen sintético construído con 6 epítopos conservados de los tres dominios de las SVMPs de *Echis ocellatus* se consiguió bloquear en un 75% el área hemorrágica inducida por el veneno completo de la misma especie. Anticuerpos frente a los epítopos 1-4 del gen sintético presentaron reactividad cruzada con metaloproteasas de *Cerastes cerastes cerastes* y, consecuentemente, los animales transfectados con el gen sintético presentaron áreas de hemorragia un 43% menores que los controles en respuesta a la inoculación de veneno completo de *Cerastes cerastes cerastes*. La idea, sin embargo, es seleccionar, mediante combinación de datos de proteómica y transcriptómica, aquellos epítopos candidatos que estén presentes en las toxinas del veneno.

Para la puesta a punto de una estrategia para la caracterización proteómica de venenos de serpientes utilizamos el veneno del crótalo *Sistrurus miliarius barbouri* por estar disponible en el laboratorio y no existir en el banco de proteínas UniProtKB/Swiss-Prot (http://us.expasy.org/sprot) más que la secuencia completa de una proteína, la disintegrina barbourina [P22827]. Además, comprobamos mediante separación de proteínas por electroforesis bidimensional (2D-SDS-PAGE) y cromatografía de fase inversa que poseía una complejidad relativamente baja. Análisis posteriores de los proteómas de otros venenos de Viperidae (Bazaa et al. 2005; Juárez et al. 2006a; Sanz et al. 2006a; Sanz et al. 2006b) mostraron que la elección de otro veneno no hubiera alterado las conclusiones generales del trabajo. El veneno analizado era una mezcla de los venenos de animales adultos de

diversas edades y de ambos sexos, no habiéndose hecho pues ningún intento de establecer posibles diferencias intraespecíficas debidas a factores individuales (edad, sexo) o medioambientales (ecosistema, dieta), como las que se han descrito en *Bothrops jararaca* (Menezes et al. 2006).

Mediante electroforesis bidimensional, las proteínas del veneno de *S.m. barbouri* se separaron en tres conjuntos mayoritarios de manchas. Dos grupos mostraban puntos isoeléctricos ácidos y masas moleculares comprendidas entre 46-14 kDa, mientras que el otro grupo mostraba un punto isoeléctrico básico y una masa aparente de 12 kDa. El resto de las proteínas migraban a lo largo del gradiente de pH (3-10) y presentaban masas moleculares que rondaban los 30-35 kDa o por debajo de 10 kDa. Para caracterizar estas proteínas, el veneno se fraccionó por cromatografía de fase inversa (RP-HPLC) obteniendose 15 fracciones mayoritarias. Cada una de estas fracciones fue analizada por SDS-PAGE, secuenciación N-terminal, determinación de masa molecular y huella peptídica por espectrometría de masas MALDI-TOF y contenido en cisteínas libres y enlaces disulfuro. Con la excepción del pico 15 (y las fracciones 1-3 que no contenían material proteico), las demás fracciones contenían proteínas con secuencias N-terminales definidas, que permitieron asignarlas a familias de proteínas conocidas. La comparación de las huellas trípticas de las proteínas separadas por 2D-SDS-PAGE y RP-HPLC permitió correlacionar bandas electroforéticas con picos cromatográficos y mostró además la existencia de múltiples isoformas de serinproteasas y de Zn^{2+}-metaloproteasas de clase PI que eluyen, sin embargo, en sólo 4 fracciones cromatográficas. La presencia de isoformas en fracciones cromatográficas aparentemente homogéneas fue posteriormente corroborada en un análisis más exhaustivo mediante separación electroforética de los componentes de cada fracción cromatográfica seguido de digestión tríptica de las bandas y secuenciación de los iones peptídicos por espectrometría de masas en tándem (MS/MS) (Sanz et al. 2006b).

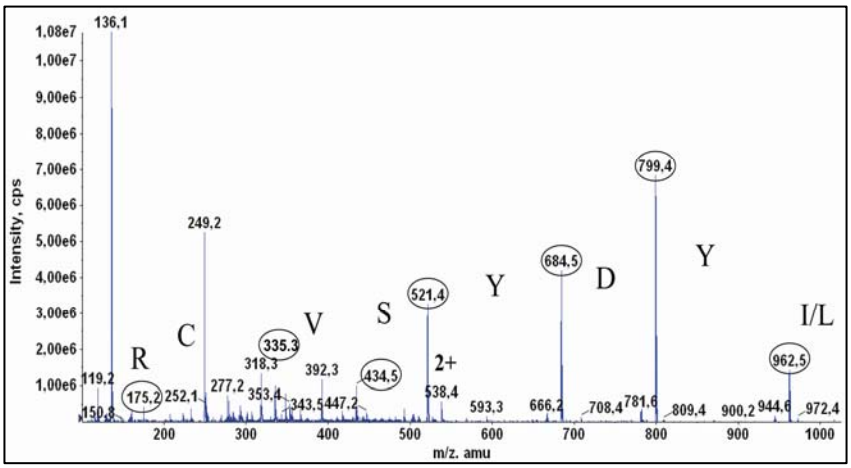
La determinación del número de cisteínas libres y de enlaces disulfuro se realiza de manera muy sencilla y con gran precisión mediante espectrometría de masas MALDI-TOF y aporta una característica casi específica de cada familia de proteínas presentes en los venenos de serpientes (Fig.14).

Masa Molecular (kDa)	Familia proteica	Nº Cisteínas
4-5	-Disintegrin corta	8
6-7	-Inhibitor tipo Kunitz	6
	-Disintegrina media	12
13-15	-PLA_2	14
	-Disintegrina dimérica	(2x10)
	-Cistatina	4
23-33	-CRISP	16
	-PI-SVMP	9
	-Fragmento DC	26
	-Serinproteasa	12
	-αβ C-lectin	(2x7)
	-sv-VEGF	(2x9)
46-58	-sv-VEGF	(2x16)
	-PIII-SVMP	37
	-LAO	6

Figura 14. Ejemplo de caracterización teórica de familias de proteínas presentes en venenos de serpientes Viperidae en base a la masa molecular y el número de cisteínas de la molécula.

Así, si bien una proteína de masa molecular 13-15 kDa pudiera corresponder a una PLA_2, una disintegrina dimérica, o una cistatina, la determinación de su contenido en cisteínas no deja lugar a dudas sobre su identidad. De manera similar, la fracción 15 de RP-HPLC (48.5 kDa) contenía un total de 35 cisteínas, de las que 1 estaba reducida y las restantes 34 formaban 17 enlaces disulfuro. Estas características son propias de una metalloproteasa de clase PIII formada por un dominio de metaloproteasa (~23

kDa, 1 cisteína libre y 4 enlaces disulfuro), un dominio tipo disintegrina (~12 kDa, 8 enlaces disulfuro) y un dominio rico en cisteínas de ~13 kDa tramado por 6 enlaces disulfuro. La asignación de la fracción 15 como una PIII-SVMP se corroboró mediante análisis de huella peptídica y secuenciación de iones trípticos por MS/MS. En particular, se caracterizaron iones correspondientes a regiones de los tres dominios: metaloproteasa (m/z 2208.04, YLYMHVALVGLEIWSNGDK), disintegrina (m/z 526.7^{2+}, GNYYGYCR) y rico en cisteínas (m/z 776.1^{2+}, VCSNGHCVDVATAY). Las secuencias de estos fragmentos trípticos están conservadas en la metaloproteasa PIII de *Bothrops jararaca* denominada jararhagina [P30431] y fueron identificadas de forma automática utilizando el programa MASCOT (http://www.matrixscience.com) y el banco de datos MSDB, de acceso público y compilado a partir de fuentes primarias como PIR, Trembl, GenBank, SwissProt y NRL3D (http://csc-fserve.hh.med.ic.ac.uk/msdb.html). Por lo general, sin embargo, las proteínas de una misma familia presentan suficiente variación de secuencia intra- e interespecífica para que sus huellas peptídicas no sean reconocibles por los motores de búsqueda que se basan en identidad de masa. La figura 15 ilustra este punto. El espectro de fragmentación del ión doblemente cargado $(M+2H)^{2+}$ a m/z 538.4 no pudo ser asignado de forma automática utilizando MASCOT, pero fue interpretado manualmente como (I/L)YDYSVCR. Una búsqueda de similitud de secuencias de vertebrados utilizando BLAST (http://www.ncbi.nlm.nih.gov/BLAST) identificó varias serinproteasas de venenos de serpientes. La mayor similitud fue con la serinproteasa-1 de *Bitis gabonica*, cuya secuencia ^{170}LFDYSVCR^{177} difiere de la homóloga de *S.m. barbouri* en un solo residuo.



> qi|38570060|gb|AAR24534.1| serine protease 1 [Bitis gabonica]
Length=260

Score = 27.4 bits (57), Expect = 14, Method: Composition-based stats.
Identities = 7/8 (87%), Positives = 8/8 (100%), Gaps = 0/8 (0%)

Query 1 LYDYSVCR 8
L+DYSVCR
Sbjct 170 LFDYSVCR 177

> qi|1351938|sp|P47797|VSP2_AGKRH Ancrod precursor (Venombin A) (Protein C activator) (ACC-C)
qi|403078|gb|AAA49195.1| ancrod
Length=258

Score = 24.8 bits (51), Expect = 73, Method: Composition-based stats.
Identities = 6/8 (75%), Positives = 8/8 (100%), Gaps = 0/8 (0%)

Query 1 LYDYSVCR 8
LY+Y+VCR
Sbjct 170 LYNYTVCR 177

> qi|6562944|emb|CAB62591.1| serine proteinase, homolog [Macrovipera lebetina]
qi|139596555|sp|Q9PT40|VSP2_VIPLE Venom serine proteinase-like protein 2 precursor
Length=260

Score = 24.4 bits (50), Expect = 96, Method: Composition-based stats.
Identities = 6/8 (75%), Positives = 8/8 (100%), Gaps = 0/8 (0%)

Query 1 LYDYSVCR 8
++DYSVCR
Sbjct 170 MFDYSVCR 177

Figura 15. Espectro de fragmentación del ión m/z 538.4^{2+} y asignación manual de la secuencia peptídica (I/L)YDYSVCR. Panel inferior, resultado de la búsqueda de similitud de secuencia utilizando BLAST.

Denominamos "venómica" a la estrategia basada en técnicas de proteómica descrita en el Artículo 1, el cual representa la primera instancia en que este acrónimo aparece en la literatura compilada en el banco de datos PubMed (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?DB=pubmed) (Fig.16).

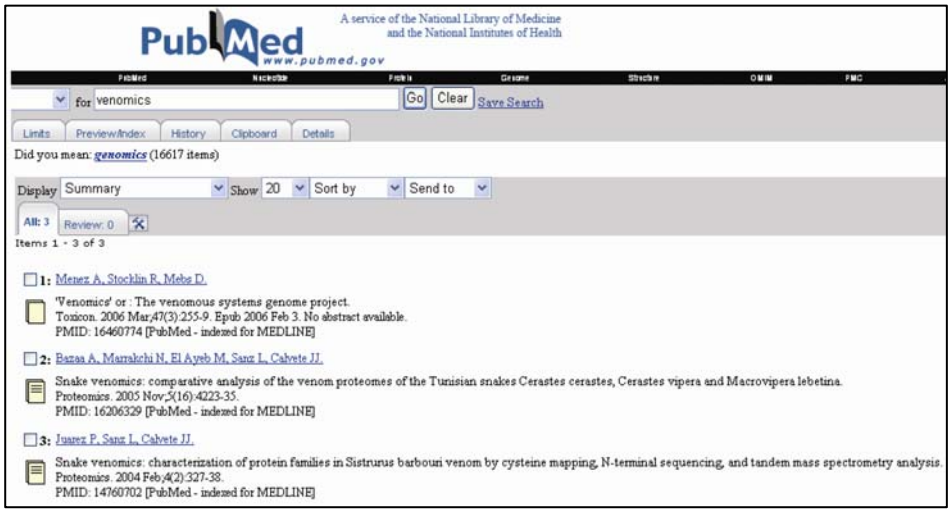


Figura 16. Resultado de la búsqueda en PubMed de artículos que incluyan la palabra "venomics". Sólo 3 artículos contienen este acrónimo.

El banco de datos Swiss-Prot/TrEMBL contiene varios cientos de secuencias (completas o parciales) de proteínas de venenos de serpientes que representan a todas las familias mayoritarias de toxinas descritas hasta la fecha. Nuestra experiencia proteómica, inicialmente con el veneno de *S.m. barbouri* y durante los últimos años con los venenos de otras especies de Viperidae de los géneros *Sistrurus*, *Crotalus*, *Agkistrodon*, *Cerastes*, *Macrovipera*, *Echis* y *Bitis*, es que a) los venenos de estas serpientes están compuestos por un pequeño número (10-12) de proteínas de diferentes familias cuya distribución y abundancia relativa varían entre especies, y b) que existe suficiente información estructural en los bancos de datos para garantizar la asignación de prácticamente todas las proteínas de un veneno dado mediante el

tratamiento automatizado o manual de datos obtenidos por espectrometría de masas.

La amplia variación en la composición porcentual de los venenos de las tres especies tratadas en esta Tesis, *S.m. barbouri* (Juárez et al. 2004; Sanz et al. 2006b), *Echis ocellatus* (Juárez et al. 2006b; Wagstaff y Harrison 2006) y *Bitis arietans* (Juárez et al. 2006a), se muestra en la Tabla 2.

	Veneno		
Familia de proteína	Smb	Ba	Eo
Disintegrina	7.7	17.8	6.8
Cistatina	-	1.7	-
BPP	< 0.1	-	< 0.1
Inhibidor Kunitz	0.1	4.1	-
Fragmento DC	1.3	-	1.7
svNGF/svVEGF	< 0.1	-	-
PLA_2	32.4	4.3	12.6
CRISP	2.9	-	1.5
Serinproteasa	17.1	19.5	2.0
C-lectina	< 0.1	13.2	7.0
L-amino acido oxidasa	2.1	-	1.4
Zn^{2+}-metalloproteasa	36.1	38.5	67.1

BPP, bradykinin-potenciating peptide; svNGF/svVEGF, snake venom nerve growth factor/vascular endothelial growth factor; CRISP, cysteine-rich secretory protein.

Tabla 2. Porcentajes de proteínas de las familias más abundantes en los venenos de serpientes estudiados. (Smb: *Sistrurus m. barbouri*; Ba: *Bitis arietans*; Eo: *Echis ocellatus*)

Los venenos de estas tres especies no solo varían en la clase y abundancia relativa de sus toxinas sino en el tipo de disintegrinas presentes. *S.m. barbouri* y *Bitis arietans* contienen únicamente disintegrinas medias (isoformas de barbourina) (Scarborough et al. 1991) o largas (isoformas de bitistatina) (Shebuski et al. 1989), respectivamente, mientras que *Echis ocellatus* secreta la disintegrina corta ocelatusina (Smith et al. 2002) y disintegrinas diméricas (EO4, EO5) (Calvete et al. 2003). Decidimos, pues, analizar librerías de cDNA de las glándulas del veneno de *Bitis arietans* y *Echis ocellatus*, construídas en el laboratorio de Rob Harrison con participación de P. Juárez, con el objetivo de comparar proteoma y transcriptoma e investigar la posible presencia de intermediarios evolutivos de las familias de disintegrinas no expresados en los venenos.

4.2. TRANSCRIPTÓMICA: CARACTERIZACIÓN DE cDNAs DE DISINTEGRINAS

4.2.1. *BITIS ARIETANS*: BA-5A, UN INTERMEDIARIO PIII-PII

Utilizando cebadores complementarios de regiones 5´y 3´no codificantes muy conservadas en secuencias de cDNA de otras disintegrinas amplificamos y secuenciamos un transcrito, denominado BA-5A, que codifica un precursor de disintegrina no expresada en el veneno de *Bitis arietans*. La secuencia deducida de la pauta abierta de lectura contiene un péptido señal, un pro-péptido, un dominio metaloproteasa y un dominio tipo disintegrina. Esta estructura modular es típica de las SVMPs de tipo PII (Fig.5). Sin embargo, un análisis más detallado del dominio tipo disintegrina de BA-5A muestra que éste posee características sólo encontradas hasta la fecha en los dominios tipo disintegrina de las SVMPs PIII, tales como presentar un motivo ECD en lugar

de RGD (o similar), y poseer 16 cisteínas. Además, un análisis filogenético (Fig.3 del Artículo 2) muestra que BA-5A segrega claramente en el grupo de os dominios tipo disintegrina de metaloproteasas de clase PIII. El hecho de carecer de dominio rico en cisteínas indica que BA-5A debe representar una nueva clase de metaloproteasas modulares PII con dominio tipo disintegrina PIII. Nuestra hipótesis es que BA-5A es un intermediario en la ruta evolutiva de una metaloproteasa PIII a una PII como la que codifica a la disintegrina larga bitistatina (Fig. 17).

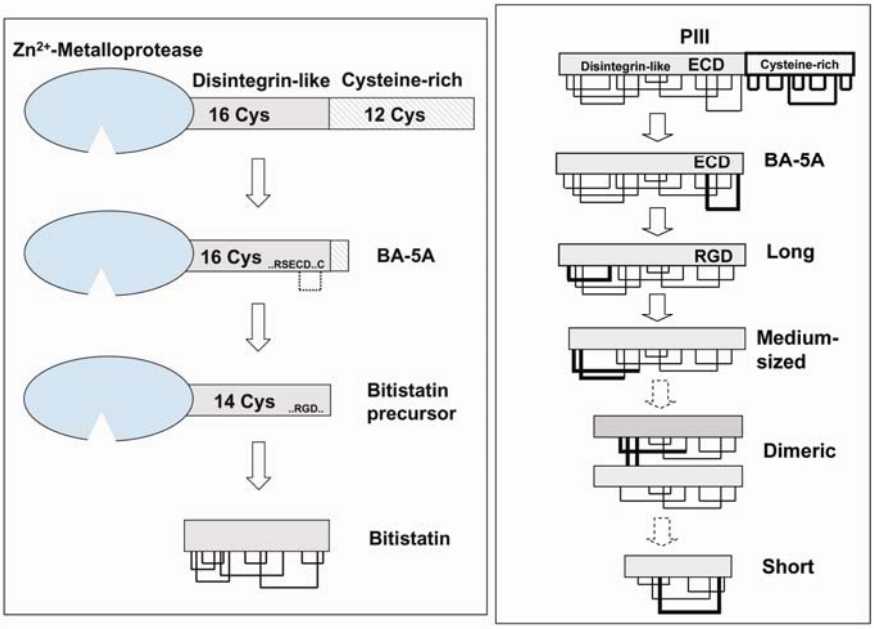


Figura 17. Esquema de la hipotética participación de BA-5A como intermediario en la ruta evolutiva de la disintegrina larga bitistatina partiendo de un precursor SVMP de clase PIII. Panel derecho, esta hipótesis implica que la pérdida del dominio rico en cisteínas antecedió a la pérdida del enlace disulfuro CysXIII-CysXVI característico de los dominios tipo disintegrina PIII y a la aparición del motivo RGD de inhibición de integrinas.

La hipótesis de que BA-5A sea un intermediario en la ruta evolutiva de una disintegrina larga, posiblemente bitistatina, implica que la deleción del dominio rico en cisteínas ocurrió con anterioridad a la pérdida del enlace disulfuro entre las cisteínas XIII y XVI, absolutamente conservado en todos los dominios tipo disintegrina PIII. Además, puesto que la cisteína XIII es el resíduo central del motivo ECD de BA-5A (también presente en muchos otros dominios tipo disintegrina PIII), la pérdida del enlace CysXIII-CysXVI por fuerza debió anteceder a la aparición del motivo RGD de inhibición de integrinas. Nuestra hipótesis es que la desaparición de este enlace disulfuro liberó una restricción estructural que posibilitó la aparición de los motivos de inhibición de integrinas en el ápice de un bucle móvil. Precisamente, se ha descrito que los bucles móviles participan en el reconocimiento rápido entre proteínas, especialmente receptores e inhibidores (Burgen et al. 1975; Williams 1989).

4.2.2. *ECHIS OCELLATUS*

4.2.2.1. DISINTEGRINAS DIMÉRICAS

En el caso de *Echis ocellatus* (Eo) amplificamos 5 precursores de disintegrinas diméricas (Fig. 18) que muestran una gran similitud de secuencia con disintegrinas diméricas descritas previamente en el veneno de esta serpiente (Calvete et al. 2003), incluyendo el número y la posición de los residuos de cisteínas.

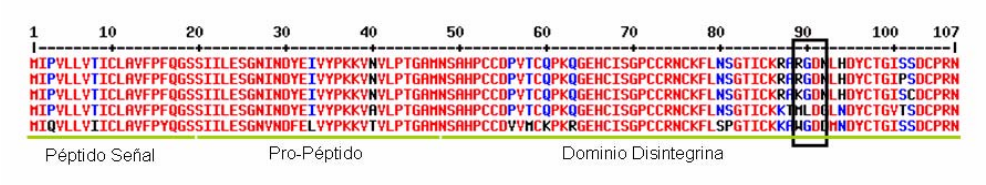


Figura 18. Alineamiento de las secuencias de aminoácidos codificados por los mensajeros de disintegrinas diméricas amplificados de la librería de cDNA de *Echis ocellatus*. Los motivos de unión a integrinas (RGD, MLD ,WGD y KGD) están encuadrados.

Todos estos clones pertenecen a la clase de "mensajeros cortos", descritos por primera vez por Okuda y colaboradores en *Agkistrodon contortrix contortrix*, denominados así porque carecen de dominio metaloproteasa. Desde entonces, este tipo de precursores de subunidades de disintegrinas diméricas se han descrito en *Bitis gabonica* (Francischetti et al. 2004), *Macrovipera lebetina transmediterranea* y *Cerastes vipera* (Sanz et al. 2006a), y va afianzándose la hipótesis de que este tipo de mRNA cortos constituyen la estructura canónica de los mensajeros de disintegrinas diméricas. Es posible que la pérdida del voluminoso dominio de metaloproteasa haya desempeñado un papel clave en la emergencia del mecanismo de dimerización de disintegrinas (Fig. 19).

Entre los amplicones de *Echis ocellatus*, los que presentan secuencias KGD y WGD representan nuevos mensajeros para disintegrinas diméricas o productos que no han podido ser detectados en el veneno. La mayor diversidad a nivel de transcriptoma que de proteoma puede indicar la existencia de un mecanismo de regulación de la expresión de estas disintegrinas, o un patrón de expresión temporal a lo largo de la vida de la serpiente, o simplemente un nivel de expresión inferior al detectable con las técnicas proteómicas empleadas.

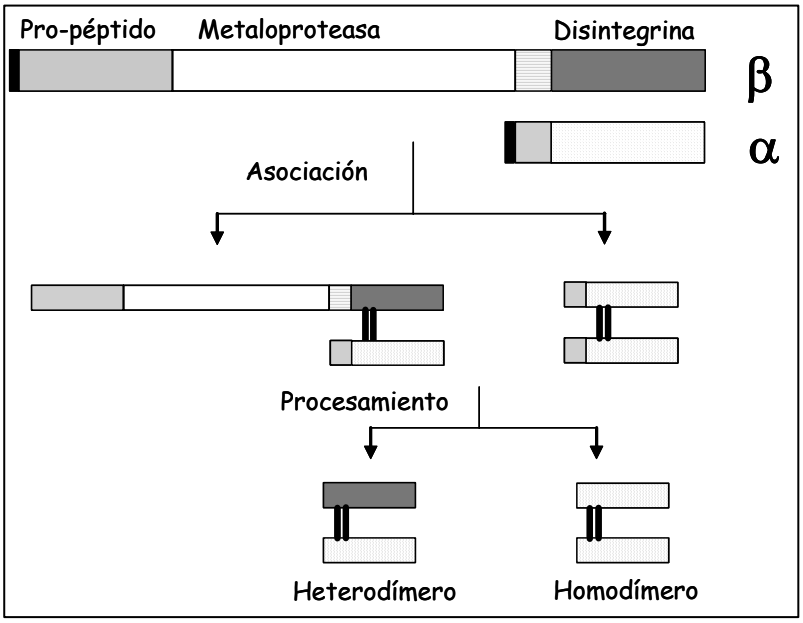


Figura 19. La existencia de precursores α y β de subunidades de disintegrinas diméricas sugiere un mecanismo de regulación de la dimerización por el cual los homodímeros resultarían de la asociación de dos precursores de subunidades tipo α, estando la asociación de dos subunidades tipo β impedida debido a impedimentos estéricos ocasionados por el gran tamaño del N-terminal de dichos precursores. Por otro lado los heterodímeros estarían formados por asociaciones de subunidades tipo α/α o bien α/β. Hasta la fecha todas las disintegrinas diméricas investigadas cumplen estas reglas empíricas.

4.2.2.2. DOS PRECURSORES DE LA DISINTEGRINA CORTA OCELLATUSIN

Además de los cDNAs de disintegrinas diméricas, amplificamos dos mensajeros (Eo-00006 y Eo10c-10) cuyas pautas abiertas de lectura codifican ambas para la misma disintegrina corta, ocelatusina (Smith et al. 2002). La diferencia entre ambos precursores reside en la presencia de un dominio metaloproteasa en el clon Eo-00006 (precursor largo) ausente en Eo10c-10 (precursor corto) (Fig.20). El precursor largo presenta pues la estructura

típica de una SVMP de clase PII mientras que el precursor corto debe clasificarse en el grupo de "mensajeros cortos" descritos en el apartado anterior.

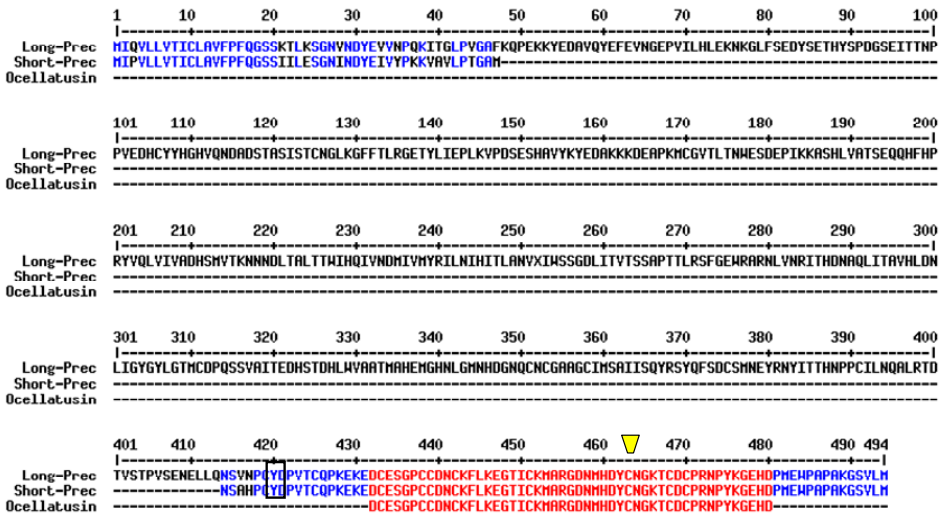


Figura 20. Alineamiento de las secuencias de la disintegrina corta ocelatusina (Smith et al. 2002) y de sus precursores largo (clon Eo-00006) y corto (Eo10c-10) (Juárez et al. 2006b). El cono amarillo señala la cisteína específica de las disintegrinas cortas y en la caja se muestra la mutación C→Y.

Un análisis filogenético de los dominios disintegrina de los precursores de ocelatusina muestra que ambos se agrupan en la misma subfamilia que las subunidades de disintegrinas diméricas (Fig.6 del Artículo 3) sugiriendo que la disintegrina corta ocelatusina se originó a partir de un gen de disintegrina dimérica. A este respecto cabe destacar que únicamente se precisan 2 mutaciones para convertir una subunidad de disintegrina dimérica en una disintegrina corta (Fig. 19). Una mutación T<u>G</u>T→T<u>A</u>T o T<u>G</u>C→T<u>A</u>C produce un cambio Cys→Tyr en la posición 420 del precursor largo (54 del precursor corto) y una segunda mutación T<u>C</u>T→T<u>G</u>T muta el codón de una serina conservada generando una cisteína en posición 468/102 (precursor

largo/precursor corto). Las consecuencias estructurales de estos dos cambios aminoacídicos son la clave para comprender la transición disintegrina dimérica→disintegrina corta. En efecto, la mutación Cys→Tyr impide la formación de los dos enlaces disulfuro (Cys7A-Cys12B y Cys7B-Cys12A) que enlazan covalentemente las dos subunidades del dímero (Fig.21) (Bilgrami et al. 2004; Bilgrami et al. 2005). Por otra parte, la mutación Ser→Cys genera el resíduo de cisteína característico de las disintegrinas cortas (CysVII, Figs. 11 y 16) necesario para formar un enlace disulfuro con la CysIII que de otra manera quedaría desapareada tras el procesamiento proteolítico de la región N-terminal del precursor de la disintegrina corta (Figs. 5 y 7 del Artículo 3).

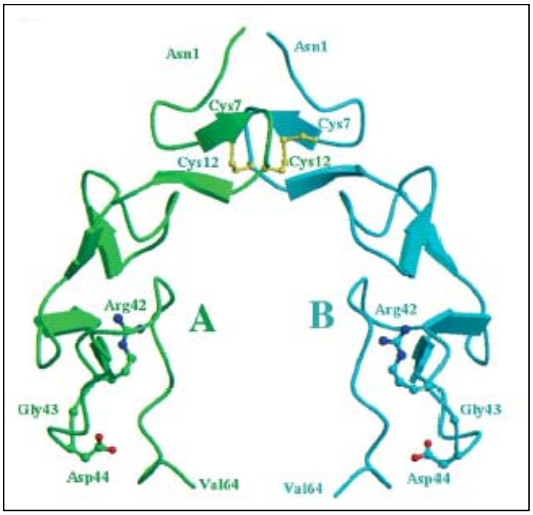


Figura 21. Estructura cristalina de la disintegrina homodimérica schistatina (Bilgrami et al. 2004). Los dos puentes disulfuro intercatenarios (Cys7A-Cys12B y Cys7B-Cys12A) se resaltan en amarillo.

Estos datos apoyan la hipótesis del origen evolutivo de ocelatusina (y posiblemente de otras disintegrinas cortas) a partir de la transformación de un precursor del grupo de las subunidades de disintegrinas diméricas, al tiempo que ponen en evidencia la gran economía molecular de esta transición.

4.3. GENÓMICA: ESTRUCTURA DE LOS GENES DE DISINTEGRINAS DIMÉRICAS Y CORTAS

La evolución biológica actúa primariamente a nivel de los genes. Buscar respuestas a preguntas sobre la diversificación estructural y funcional de la familia de las disintegrinas requiere, pues, un conocimiento detallado de la estructura de sus genes. No disponemos de la secuencia del genoma de ninguna serpiente, por lo que preguntas como el número de genes de disintegrinas, su compartimentalización, regulación de su expresión, estructura de promotores, etc. no han sido abordadas. Desde un punto de vista de la evolución de las toxinas en general, y de las disintegrinas en particular, sería extremadamente interesante investigar la presencia de pseudogenes. Un pseudogen es un segmento de ADN similar a un gen funcional, pero que contiene cambios de nucleótido que impiden su transcripción o traducción. Se cree que los pseudogenes se originan en la duplicación génica o en la transcripción inversa de ARN.

Al comienzo de esta Tesis sólo había depositada en la base de datos GenBank (http://www.ncbi.nlm.nih.gov) la secuencia de nucleótidos del gen de las isoformas 2 y 3 de la disintegrina media Halystatina (D28871) de *Gloydius halys*, y recientemente el grupo de John C. Perez (Natural Toxins Research Center, Kingsville, Texas, USA) ha publicado secuencias genómicas parciales que codifican para una región de 41 aminoácidos de diversas disintegrinas homólogas de subespecies de *Agkistrodon contortrix* (Soto et al. 2006). Nuestros datos muestran que los genes de las tres subunidades de disintegrinas diméricas de *Echis ocellatus* (Eo) (2) y *Macrovipera lebetina transmediterranea* (Mlt) (1) poseen idéntica estructura global, un intrón central de alrededor de 1000 nucleótidos (nt) flanqueado por dos exones cortos de 93-102 nt (Figuras 2 y 3 del Artículo 4). Por otra parte, es de

reseñar, que todas las secuencias genómicas amplificadas de Eo y Mlt que codifican para disintegrinas cortas carecen de intrones (Figs. 2C, 3B y 3C del Artículo 4).

Las secuencias de los intrones guardan 88-90% de identidad entre si y con el intrón 2 de de la disintegrina media Halystatina. Además, tanto el tamaño, el lugar de inserción, y las secuencias consenso para la eliminación del intrón (5´-*GT*AAG (donante)/3´-*AG* (aceptor)) están absolutamente conservados en las tres secuencias de Eo y Mlt, así como en las secuencias parciales de disintegrinas diméricas de *Agkistrodon contortrix*, en la secuencia del gen de la disintegrina media halystatina y en la estructura génica de los dominios tipo disintegrina de metaloproteasas de la familia ADAM, las cuales compartieron un antecesor común con las SVMPs de clase PIII. Aunque la estructura génica de una SVMP PIII no ha sido descrita hasta la fecha, una comparación de los lugares de inserción de intrones en dominios de disintegrina (Fig. 22) muestra que ha habido una pérdida selectiva de intrones a lo largo de la ruta evolutiva de las disintegrinas en paralelo a una minimización de la estructura proteica (Fig. 23).

Figura 22. Comparación de los lugares de inserción de intrones en dominios tipo disintegrina de metaloproteasas celulares de la familia ADAM (hu, humano; ck, pollo; mm, ratón; zf, pez cebra), en las isoformas 2 y 3 de la disintegrina media Halystatina, en la estructura parcial de subunidades de disintegrinas diméricas de *Agkistrodon contortrix* (Soto et al. 2006a) y en las subunidades de disintegrinas diméricas de *Echis ocellatus* (Eo_D3) y *Macrovipera lebetina transmediterranea* (Ml_G1 y Ml_G2). Nótese que los genes de disintegrinas cortas (Ml_G3, Eo_C3 y Eo_RTS) carecen de intrones.

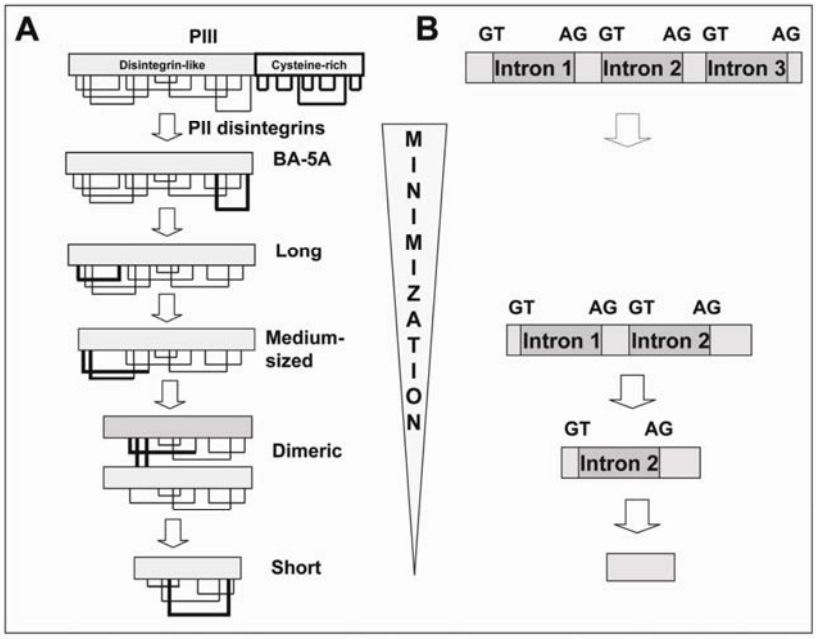


Figura 23. (A) Esquema de la ruta evolutiva de la diversificación estructural de la familia de las disintegrinas actualizado con los datos descritos en esta Tesis, incluyendo el intermediario BA-5A. El panel B muestra la organización génica del dominio tipo disintegrina de una metalopreoteasa de la familia ADAM y de las subunidades de disintegrinas diméricas y cortas de *Echis ocellatus* y *Macrovipera lebetina transmediterranea* (Artículo 4). Resulta aparente el paralelismo entre la minimización de las estructuras proteica y génica a lo largo de la evolución de la familia de las disintegrinas.

El estudio de la organización génica de los genes que codifican para disintegrinas nos ha aportado información extra a la que habiamos obtenido del estudio de los mensajeros que las codifican. Ahora podemos afirmar que el mecanismo molecular de la diversificación estructural y funcional de esta familia de proteínas engloba no sólo una sucesión de deleciones y mutaciones puntuales, sino también una pérdida de secuencias intrónicas. No obstante, un entendimiento más profundo de la ruta evolutiva propuesta, así como la generalización de que la pérdida de intrones forma parte del mecanismo evolutivo de las disintegrinas, requiere un estudio exhaustivo de las estructuras de los genes de miembros de todos los grupos de disintegrinas en diversas especies de serpientes.

5. ANEXOS

5.1. TRABAJO 1: Snake venom disintegrins: evolution of structure and function

ELSEVIER

Toxicon 45 (2005) 1063–1074

TOXICON

www.elsevier.com/locate/toxicon

Snake venom disintegrins: evolution of structure and function

Juan J. Calvete[a,*], Cezary Marcinkiewicz[b], Daniel Monleón[c], Vicent Esteve[a,c], Bernardo Celda[c,d], Paula Juárez[a], Libia Sanz[a]

[a]*Instituto de Biomedicina de Valencia, C.S.I.C., Jaume Roig 11, 46010 Valencia, Spain*
[b]*Department of Biology, Center for Neurovirology and Cancer Biology, Temple University, N. 12th Street, Philadelphia, PA 19122, USA*
[c]*Departamento de Química Física, Universitat de València, Dr Moliner 50, 46100 Burjassot (Valencia), Spain*
[d]*Servicio Central de Soporte a la Investigación Experimental, Universitat de València, Dr Moliner 50, 46100 Burjassot (Valencia), Spain*

Received 23 February 2005
Available online 12 April 2005

This paper is dedicated to the memory of Stefan Niewiarowski

Abstract

Disintegrins represent a family of polypeptides present in the venoms of various vipers that selectively block the function of integrin receptors. Here, we review our current view and hypothesis on the emergence and the structural and functional diversification of disintegrins by accelerated evolution and the selective loss of disulfide bonds of duplicated genes. Research on disintegrins is relevant for understanding the biology of viper venom toxins, but also provides information on new structural determinants involved in integrin recognition that may be useful in basic and clinical research. The role of the composition, conformation, and dynamics of the integrin inhibitory loop acting in concert with the C-terminal tail in determining the selective inhibition of integrin receptors is discussed.

Keywords : Snake venom proteins; Disintegrins; Integrin antagonists; Structure-function correlations; Evolution of protein structure; Disulfide bond engineering

1. Introduction: snake venom toxins

Snake venoms contain complex mixtures of hundreds of pharmacologically active molecules, including organic and mineral components (histamine and other allergens, polyamines, alkaloids…), small peptides and proteins (Markland, 1998; Fry, 1999; Ménez, 2002). The biological effects of venoms are complex because different components have distinct actions and may, in addition, act in concert with other venom molecules. Toxic venom proteins play a number of adaptive roles: immobilizing, paralyzing, killing, liquefying prey and deterring competitors. The synergistic action of venom proteins may enhance their activities or contribute to the spreading of toxins. According to their major toxic effect in an envenomed animal, snake venoms may be conveniently classified as neurotoxic and haemorrhagic. Among the first group are the *Elapidae* snakes (mambas, cobras, and particularly the Australian snakes, which are well known to be the most toxic in the world). Venoms of *Viperidae* and *Crotalidae* snakes (vipers and rattlesnakes) contain a number of different proteins that interfere with the coagulation cascade, the normal haemostatic system and tissue repair. Consequently, envenomation by these snakes generally results in persistent bleeding. The proteins in haemorrhagic venoms can be grouped into a few major protein families, including enzymes (serine proteinases, Zn^{2+}-dependent PI–PIV metalloproteinases of the reprolysin family, and group

* Corresponding author. Tel.: +34 96 339 1778; fax: +34 96 369 0800.
E-mail address: jcalvete@ibv.csic.es (J.J. Calvete).

0041-0101/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.toxicon.2005.02.024

II phospholipase A_2 isoenzymes) and proteins with no enzymatic activity (C-type lectins, CRISP, and disintegrins) (Ménez, 2002; Juárez et al., 2004).

Snake venom hemorrhagic metalloproteinases (SVMP) are clustered with mammalian matrix-degrading metalloproteinases and proteins of the ADAM familiy in a monophyletic evolutionary tree (Moura da Silva et al., 1996). The monophyletic distribution of the mammalian and snake venom proteins indicate that SVMPs have evolved relatively late during evolution from a common ancestor gene both by speciation (after mammals and reptiles diverged) and by gene duplication, followed by divergence of the copies through positive Darwinian selection. SVMPs have been classified according to their domain structure into four classes (Hite et al., 1994; Jia et al., 1996). All four groups share homologous signal, pro-domains, and proteinase domains. The PI metalloproteinases (20–30 kDa) are single-domain proteins with relatively weak hemorrhagic activity. The class PII metalloproteinases (30–60 kDa) (Bjarnason and Fox, 1994) contain a disintegrin domain at the carboxyl terminus of a metalloproteinase domain structurally similar to that in the class PI. Hemorrhagins of the PIII class are large toxins (60–100 kDa) with the most potent activity, and comprise multidomain enzymes built up by an N-terminal metalloproteinase domain and C-terminal disintegrin-like and cysteine-rich domains (Jia et al., 1996; McLane et al., 1998). The PIV class of SVMPs has a similar domain structure as the PIII class, but with additional disulfide-linked C-type lectin-like domain(s). Disintegrins and disintegrin-like domains are released in the venoms by proteolytic processing of PII and PIII metalloproteinases, respectively, (Kini and Evans, 1992), and represent potent inhibitors of integrin–ligand interactions.

2. Disintegrins: evolutionary structure diversification by disulfide bond engineering

Disintegrins, a family of small (40–100 aminoacids), cysteine-rich polypeptides, were first described as potent inhibitors of the platelet fibrinogen receptor, integrin $\alpha_{IIb}\beta_3$ (Huang et al., 1987). The isolation and characterization of disintegrins that do not inhibit platelet aggregation (i.e. non-RGD-containing dimeric disintegrins) was achieved with the development in the late 1990s of cell adhesion inhibition assays using cell lines expressing defined integrins (Marcinkiewicz et al., 1999a,b). Currently, the disintegrin family can be conveniently divided into five different groups according to the length and the number of disulfide bonds of the polypeptides (Calvete et al., 2003) (Fig. 1). The first group includes short disintegrins composed of 41–51 residues and four disulfide bonds. The second group is formed by the medium-sized disintegrins which contain

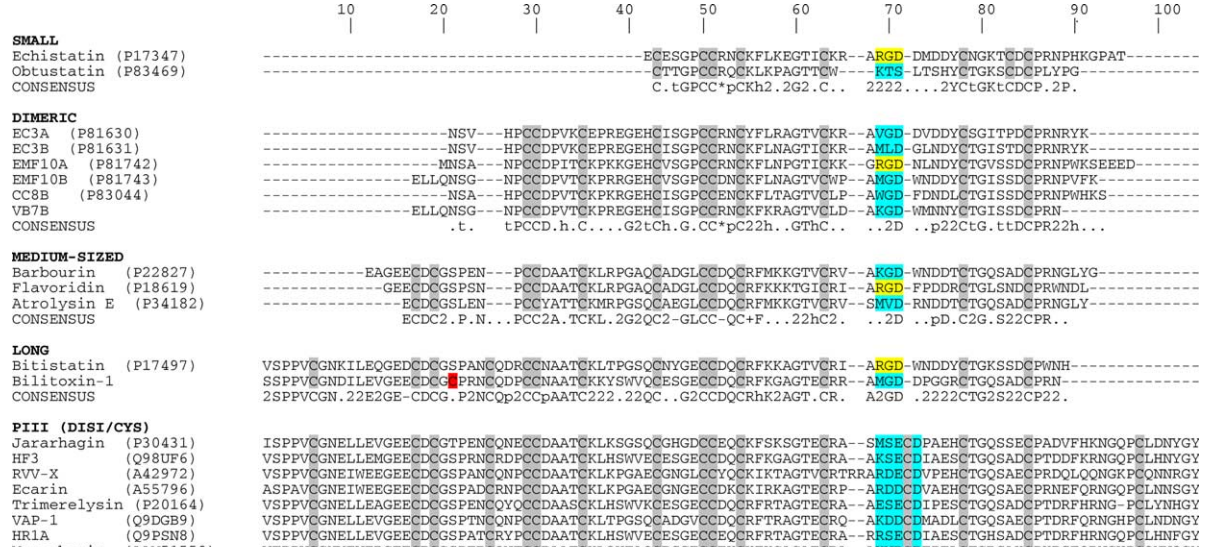


Fig. 1. Multiple sequence analysis of selected polypeptides of the different disintegrin subfamilies. Representative members of each of the five subfamilies of disintegrins displaying different integrin-binding motifs were choosen. For a more complete sequence alignment, please consult Calvete et al., 2003. The one-letter-code for amino acid nomenclature is used. Cysteine residues are shadowed in pale gray. Amino acid position where conserved cysteines are missing are highlighted in green. The extra cysteine residue of bilitoxin is colored red. The integrin-binding tripeptide arginine-glycine-aspartic acid (RGD) is depicted in yellow; non-RGD integrin-binding tripeptide motifs are shown in blue. Amino acid characteristics which define the signature of each disintegrin subfamily are shown in the 'Consensus' line using the following code: a, aromatic (F, Y, W); h, hydrophobic (L, I, V, M, A); t, turn-like or polar (G, P, N, Q, H, S, T); −, negatively charged (E, D); +, positively charged (K, R); *, charged (E, D, K, R); p, conservative (N, D, Q, E); 2, one of two residues in any sequence. When available, databank accession codes are given. The sequences of bilitoxin-1 and gramelysin-1 are from Nikai et al., 2000; Wu et al., 2001, respectively.

about 70 amino acids and six disulfide bonds. The third group includes long disintegrins with an ~84-residue polypeptide cross-linked by seven disulfide bonds. The fourth subfamily of disintegrins groups the disintegrin-like domains derived from PIII-SVMPs. PIII disintegrins are modular proteins containing an N-terminal disintegrin-like domain of about 100 amino acids including16 cysteine residues involved in the formation of eight disulfide bonds, and a C-terminal 110–120-residue cysteine-rich domain crosslinked by six disulfides (Calvete et al., 2000a). Unlike the PII (short, medium and long) and PIII disintegrins, which are single-chain molecules, the fifth group is composed of homo- and heterodimers. Dimeric disintegrins contain subunits of about 67 residues with 10 cysteines involved in the formation of four intra-chain disulfide bonds and two interchain cystine linkages (Calvete et al., 2000b; Bilgrami et al., 2004). Bilitoxin-1 represents another homodimeric disintegrin comprising disulfide-bonded polypeptides, each containing 15 cysteinyl residues (Nikai et al., 2000).

Fig. 2 displays a dendrogram for the multiple sequence analysis of the consensus sequences of disintegrin domains listed in Fig. 1. The most prominent characteristic of this tree is that the members of the different subfamilies are almost perfectly clustered separately with their homologues, suggesting an evolutionary relationship between the different disintegrin subfamilies. Disintegrins are small in size and possess a high density of disulfide bonds. A close examination of the conserved cysteine residues in each disintegrin subfamily strongly indicates that the structural diversity of disintegrins has been achieved during evolution through the selective loss of disulfide bonds (Calvete et al., 2003) (Fig. 3). PIII snake venom metalloproteinases are the closest homologues of cellular ADAMs. Comparison of full-length cDNA sequences of ADAMs and PIII SVMPs showed that the gene coding for the latter molecules possess 3′untranslated regions, which include STOP codons after the cysteine-rich domain (Fig. 4). Thus, PIII SVMPs are not simply derived by posttranslational proteolysis of ADAM molecules, but have rather evolved from a common

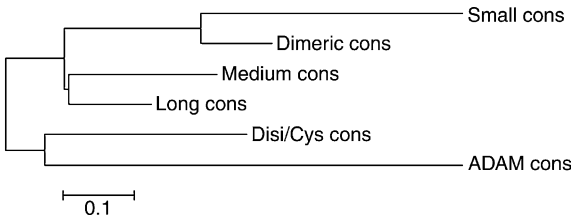


Fig. 2. Dendrogram for the multiple consensus sequence analysis of disintegrin domains. The dendogram includes also the disintegrin-like domains of ADAMs from which snake venom disintegrins evolved after mammals and reptiles diverged. The tree represents the minimum evolutionary distance estimated through neighbour joining using maximum likelihood distances. Maximum parsimony produced a similar topology. The length of the horizontal scale bar represents 10% divergence.

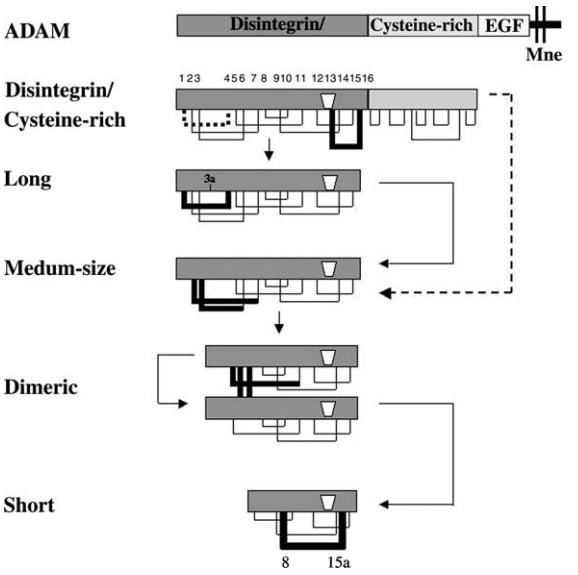


Fig. 3. Scheme of the domain organization, disulfide bonds and proposed evolutionary relationship between the different disintegrin subfamilies. A precursor ADAM molecule is shown at the top. Cysteines are numbered from 1–16. Disulfide bonds, which are removed in the hypothetical evolutionary pathway from Disintegrin/Cysteine-rich proteins to short disintegrins are displayed in thick lines, and possible evolutionary pathways in the evolution of the different disintegrin subfamilies are indicated by arrows. The disulfide bond Cys1–Cys4 absent in Graminelysin, a member of the Disintegrin/Cysteine-rich subfamily, is shown in dotted line. A hypothetical pathway from a gramelysin-type disintegrin to a medium-sized disintegrin is indicated by a dotted arrow. The cysteine residues engaged in the short disintegrin-specific disulfide bond are labeled. 3a indicates the position of an extra cysteine in bilitoxin-1. The position of the integrin binding motif is shown as a white cone.

ancestor after having lost the genetic information coding for protein regions downstream of the cysteine-rich domain (EGF-like, transmembrane and cytoplasmic domains). Further deletions of gene regions coding for the C-terminal portion of the disintegrin-like domain (including cysteine-16 and the cysteine-rich domain) and for cysteine-13, which is disulfide-bonded to Cys-16, gave rise to long disintegrins (Fig. 4). Mutations in the codon of Cys-1 along with a deletion of nine nucleotides coding for the tripeptide CQ(D/N) from a long disintegrin ancestor resulted in removal of the disulfide bond between cysteines-1 and -4 and the emergence of medium-sized disintegrins. The C1–C4 and C13–C16 disulfide bonds impose structural constraints and create loop structures that may contain functional epitopes. This is particularly evident in the case of the PIII modular disintegrins, where the putative integrin recognition loop is disulfide-bonded to the C-terminal region. Removal of this constraint may have act as a driving force in the subsequent evolution of the integrin-binding loop of the PII disintegrins.

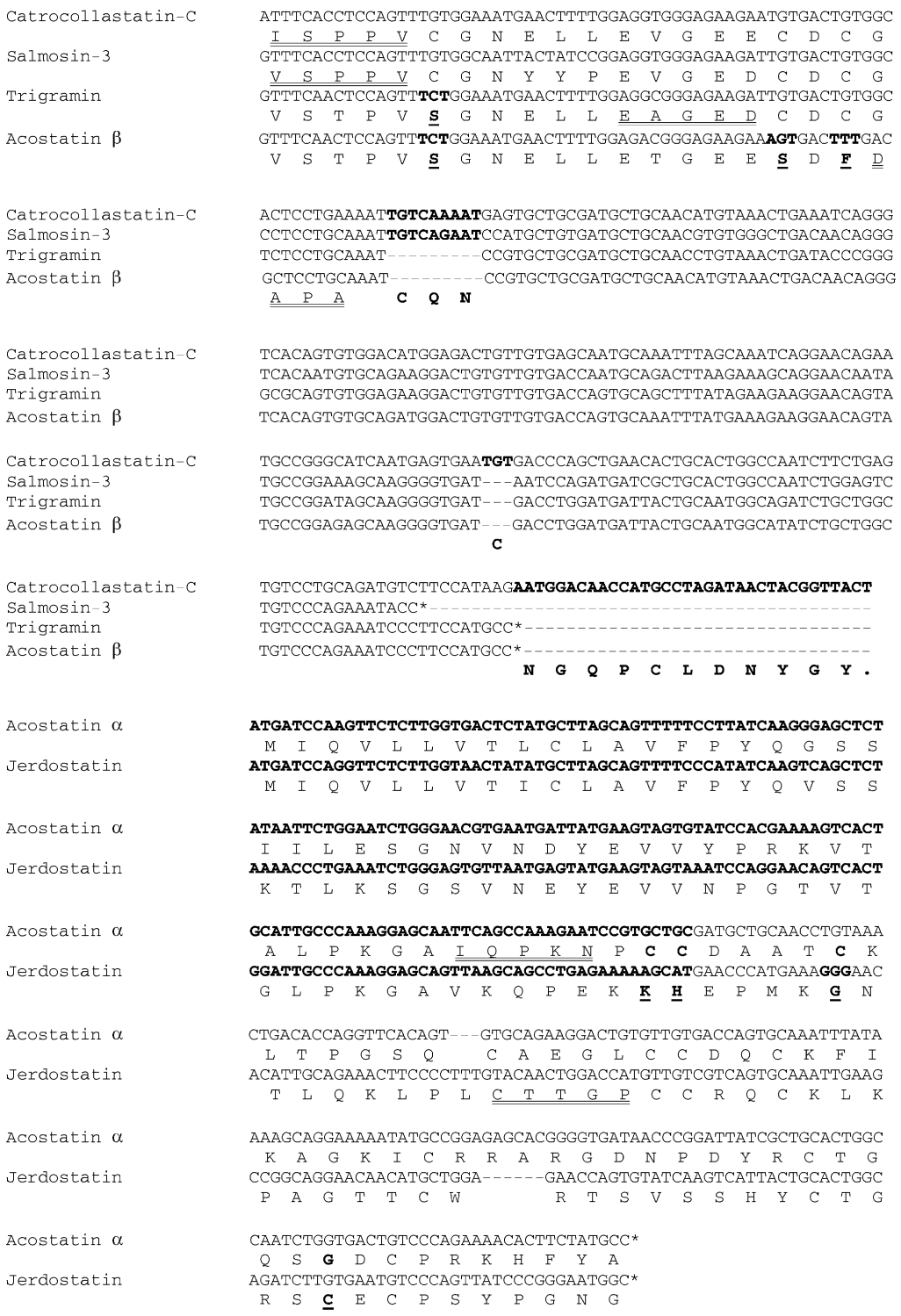


Fig. 4. Comparison of cDNA and deduced amino acid sequences. Alignment of the nucleotide sequences of cDNAs coding for representative members of the disintegrin subfamilies, catrocollastatin-C (AAC59672), (disintegrin-like/cysteine-rich), salmosin-3 (AR287876) (long), trigramin (X51530) (medium-sized), and acostatin α (AB078903) and β (AB078904) chains (dimeric), and jerdostatin (AY262730) (short). Mutations and deletions affecting the cysteine residue content in the molecules are in bold. N-terminal sequences of the mature proteins are double underlined.

Mutations involving the codons of the first two cysteine residues of medium-size disintegrins (Cys-2 and Cys-3) yielded polypeptides with 10 cysteines (Fig. 4). The disulfide bond pattern of these disintegrin chains follows the same scheme as that of medium-sized disintegrins (Fig. 3), except that the two cysteines (Cys-6 and Cys-7), which in medium-sized disintegrins are disulfide-bonded to Cys-3 and Cys-2, respectively, are engaged in two interchain disulfide bonds from another 10-cysteine-containing disintegrin in either a parallel (Cys6A–Cys6B and Cys7A–Cys7B in EMF-10; Calvete et al., 2000b) or an antiparallel arrangement (Cys6A–Cys7B and Cys7A–Cys6B in Schistatin; Bilgrami et al., 2004) giving rise to homo- and heterodimers. The fact that the amino acid sequences of the subunits of the dimeric disintegrins piscivostatin and contortrostatin cluster with the medium-sized disintegrins (Fig. 4 in Calvete et al., 2003) further suggests that these dimeric disintegrins may have evolved from a medium-size disintegrin ancestor.

Noteworthy, the cDNAs coding for the subunits of the dimeric disintegrins acostatin (*Agkistrodon contortrix contortrix*) and piscivostatin (*Agkistrodon piscivorus piscivorus*) have quite different lengths (Okuda et al., 2002). The β-subunit precursors of acostatin and piscivostatin are both synthesized from canonical protein-coding regions of about 2000 bp encoding pre-peptide, metalloproteinase, and disintegrin domains. However, the precursors of the α-subunits of both dimeric disintegrins have short (∼900 bp) coding regions and the deduced amino acid sequences consist of a signal peptide, a 30-residue pre-peptide domain (almost identical to the N-terminal portion of the β-chain pre-peptide) and the disintegrin domain (Fig. 4). Hence, the α-chain precursor genes may have evolved from a duplicated canonical disintegrin precursor gene by deletion of a continuous ∼1100 bp ORF that in the β-subunit genes encodes the C-terminal part of the pre-peptide, the metalloproteinase domain, and the disintegrin domain N-terminal region, which in the long and medium-sized disintegrins contains the cysteine residues 1–4 (Figs. 3 and 4; see also Fig. 6 in Okuda et al., 2002). Homo- and heterodimeric disintegrins have been reported in a number of *Viperidae* snakes and restricted combinations of both dimeric arrangements (which in many cases share a subunit) often co-exist in the same venom (Marcinkiewicz et al., 1999a, 2000; Calvete et al., 2002, 2003). The occurrence of precursors of the α and the β subunits of dimeric disintegrins differing in their domain structure immediately suggests a mechanism for regulation of dimerization (Fig. 5). Homodimers may result from disulfide bonding of two α-type subunit precursors, whereas the association of two β-type subunit precursors might be sterically hindered by the large N-terminal multidomain structure. On the other hand, heterodimers may be generated by α-type/α-type and α-type/β-type subunit associations.

The cDNA sequence AY262730 (http://www.ebi.ac.uk) of the non-RGD-containing short disintegrin jerdostatin

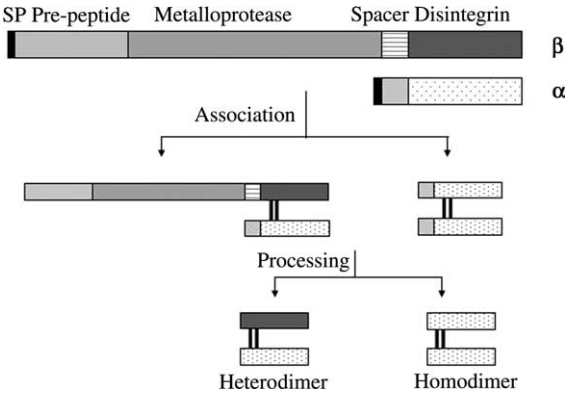


Fig. 5. Proposed mechanism for the generation of dimeric disintegrins. The precursors of the α and the β subunits of dimeric disintegrins are synthesized from genes differing in their domain structure. The β-subunit precursors contain signal peptide (SP), pre-peptide, metalloprotease, and disintegrin domains. The precursors of the α-subunits consist of a SP, a 30-residue pre-peptide domain and the disintegrin domain. Homodimers may result from disulfide bonding of two α-type subunit precursors, whereas the association of two β-type subunit precursors might be sterically hindered by the large N-terminal multidomain structure. On the other hand, heterodimers may be generated by α-type/α-type and α-type/β-type subunit associations.

(*Trimeresusus jerdonii*) shows a short coding region, which is highly homologous to that of the α-type subunit precursor of dimeric disintegrins. This strongly suggests that short disintegrins and the α-type subunit precursor of dimeric disintegrins may have a common ancestry (Fig. 3). Comparison of the cDNA sequences of acostatin-α and jerdostatin (Fig. 4) indicates that the emergence of the short disintegrins occurred by mutations in the codons of the three N-terminal cysteines of an α-type disintegrin subunit precursor and the appearance of another cysteine residue between Cys-15 and Cys-16 (15a in Fig. 3), involved in the formation of a short-disintegrin-specific disulfide bond with the otherwise unpaired Cys-8 (Fig. 3).

Though the vast majority of disintegrins might follow the canonical scheme outlined above, the evolutionary scenario of the disintegrin family might be more complex. Thus, graminelysin departs from this pathway. It contains the Cys13–Cys16 disulfide bond specific of disintegrin/cysteine-rich domains but clusters with the medium-sized disintegrins, and like them has Ser in place of Cys-1 and lacks the CQ(D/N) region. Hence, graminelysin might represent an intermediate step in an alternative route in the evolution of medium-sized disintegrins from Disintegrin-like/Cysteine-rich proteins. Bilitoxin-1, a long disintegrin from *Agkistrodon bilineatus* venom (Nikai et al., 2000), possesses an extra cysteine residue between cysteines 3 and 4 (labeled as 3a in Fig. 3) engaged in the formation of a disulfide-bonded homodimer. In addition, the disintegrins albolabrin (from *Trimeresusrus albolabris*; Calvete et al., 1991), saxatilin (isolated from the venom of the Korean

snake *Glydius saxatilis* by Hong et al., 2002), and salmosin from *Agkistrodon halys* venom (Shin et al., 2003) have the conserved cysteine residue pattern of the medium-sized disintegrin group but exhibit a distinct disulfide bond arrangement (C1–C3, C2–C4, C5–C8, C6–C11, C7–C9, C10–C12). Studies with recombinant protein species produced by site-directed mutagenesis of saxatilin evidenced that the disulfide bonds involving the N-terminal cysteine residues 1–4 do not appear to be critical for the molecular structure and biological function (inhibition of platelet aggregation) of saxatilin, whereas deletion of any of the disulfide bonds between C5–C8, C6–C11, and C10–C12 abolished the biological activity of the disintegrin. These results strongly support our evolutionary scheme outlined in Fig. 3 based on structure diversification through the sequential loss of N-terminal disulfide linkages.

Whether the non-canonical disulfide bond pattern of albolabrin/saxatilin/salmosin is restricted to a few disintegrin molecules or is more widely distributed in monomeric and dimeric disintegrins remains an open question.

3. The integrin-binding loop: functional diversification

The PIII modular disintegrins are the closest homologues of ADAMs, and represent therefore the most ancient group of the disintegrin family. The disintegrin-like domains of PIII SVMPs molecules lack the RGD cell-binding motif present in the integrin-binding loops of many PII disintegrins (see below), which is replaced by XXCD sequences. The recombinant disintegrin-like/cysteine-rich domains of *Crotalus atrox* hemorrhagin atrolysin A, as well as synthetic peptides containing the sequence RSECD, inhibit collagen-induced platelet aggregation (i.e. inhibition of integrin $\alpha_2\beta_1$) (Jia et al.,1997; Markland, 1998). For these peptides to have significant inhibitory activity, the cysteine residue must be constrained by participation in a disulfide bond. Additionally, the two acidic amino acids flanking the cysteine residue have been found to be important for biological activity (Jia et al.,1997). Similarly, jararhagin-C and catrocollastatin-C, two MSECD-containing disintegrin-like and cysteine-rich domain fragments naturally derived, respectively, from PIII hemorrhagic toxins from *Bothrops jararaca* and *Crotalus atrox*, inhibit collagen- and ADP-induced (catrocollastatin-C) by blocking platelet interaction with collagen (Zhou et al., 1995; Shimokawa et al., 1997). Jararhagin, and the highly-related MSECD-containing PIII metalloproteinase alternagin from *Bothrops alternatus*, specifically interfere with the binding of fibrillar type-I collagen to $\alpha_2\beta_1$-expressing cells (Souza et al., 2000; Moura da Silva et al., 2001; Zigrino et al., 2002). However, the involvement of the disintegrin-like domain in integrin recognition is controversial. For example, Ivaska et al. (1999) have reported that synthetic peptides displaying the RSECD sequence of the disintegrin-like domain of jararhagin failed to inhibit recombinant α2 I-domain binding to collagen, whereas a basic peptide (RKKH) from the metalloproteinase domain proved to be functional. Thus, peptides from the α2 I-domain containing the RKKH sequence bound in a Mg^{2+}-dependent manner near the MIDAS site where a collagen(I)-derived peptide is known to bind. In addition, a jararhagin-derived RKKH-peptide induced structural changes in, and competitively inhibit the binding of collagen I to, the α1 I-domain of the human $\alpha_1\beta_1$ integrin (Nymalm et al., 2004).

The function of the cysteine-rich domain is also essentially unknown, although there is evidence that the cysteine-rich domains of atrolysin A and jararhagin contain sequences that antagonize the function of the $\alpha_2\beta_1$ integrin on platelets, MG-63 and α_2-transfected K562 cells (Jia et al., 2000; Kamiguti et al., 2003).

The fact that peptides derived from all three domains of PIII metalloproteinases can inhibit the collagen-binding integrin $\alpha_2\beta_1$, indicate that SVMPs might contain several integrin recognition sequences. These sites may serve to target full-length haemorrhagic toxins to a particular site of action leading to the degradation of platelet β_1 integrin receptors (Kamiguti et al., 1995) and extracellular matrix components, which result in local and systemic hemorrhage in viperid and crotalid envenoming (Fox and Long, 1998). However, it is not clear how the metalloproteinase, disintegrin-like and cysteine-rich domains of SVMPs are spatially arranged, and how the molecular architecture of these multidomain proteins affects the protein binding capabilities of individual domains. We hypothesize that the concerted loss of the C13–C16 integrin loop-constraining linkage and the cysteine-rich domain paved the way for the emergence of single-domain PII disintegrins with more versatile integrin recognition loops.

PII disintegrins block the function of integrin receptors with a high degree of selectivity. Hence, disintegrins have found numerous applications in studies on a variety of biological processes in which integrins play pivotal roles (Niewiarowski et al., 2002). Selective blockade of integrins is a desirable goal for the therapy of a number of pathological conditions, including acute coronary ischaemia and thrombosis ($\alpha_{IIb}\beta_3$), tumor metastasis, osteoporosis, restenosis, rheumatoid arthritis ($\alpha_v\beta_3$), bacterial infection, vascular diseases ($\alpha_5\beta_1$), inflammation, autoimmune diseases ($\alpha_4\beta_1$, $\alpha_7\beta_1$, $\alpha_9\beta_1$), and tumor angiogenesis ($\alpha_1\beta_1$, $\alpha_v\beta_3$). The relevant integrin receptors involved in the above listed pathologies are among the targets of many disintegrins (Fig. 6).

The integrin inhibitory activity of PII disintegrins depends on the appropriate pairing of cysteines, which determine the conformation of the inhibitory loop. In most single-chain PII disintegrins the active sequence is the tripeptide RGD (McLane et al., 1998), the exceptions being barbourin (Scarborough et al., 1991) and ussuristatin 2 (Oshikawa and Terada, 1999), two medium-sized disintegrins possessing an active KGD sequence, and atrolysin E (Hite et al., 1992), which has an MVD motif in its inhibitory loop. RGD-containing disintegrins show different binding

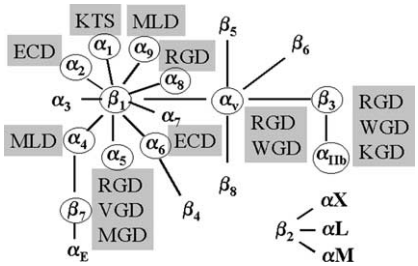


Fig. 6. Diagram of the integrin family and the different disintegrin tripeptide motifs that block specific integrin–ligand interactions. Integrin heterodimers antagonized by snake venom disintegrins are encircled. Integrins are a major class of cell surface transmembrane type I receptors that mediate a wide variety of cell–cell and cell-extracellular matrix interactions (Hynes, 2002). The integrin α and the β subunits combine in a restricted manner to form different dimers, each of which exhibits a distinct ligand-binding profile. $\alpha_1\beta_1$ is a receptor for collagen IV; $\alpha_2\beta_1$ binds collagen I; $\alpha_4\beta_1$ interacts with fibronectin and VCAM-1; $\alpha_4\beta_7$ bind the same ligands as $\alpha_4\beta_1$ and in addition is a receptor for MdCAM; $\alpha_5\beta_1$ represents the major fibronectin receptor; integrin $\alpha_6\beta_1$ is a laminin receptor; $\alpha_8\beta_1$ and $\alpha_9\beta_1$ bind tenascin; $\alpha_v\beta_1$ and $\alpha_v\beta_3$ are major vitronectin receptors; and $\alpha_{IIb}\beta_3$ is the platelet fibrinogen receptor involved in platelet aggregation. For a detailed description of the integrin family and the ligands of the different heterodimers, consult The Integrin Page at http://integrins.hypermart.net/.

affinity and selectivity towards integrins which recognise the RGD sequence in their ligands (i.e. $\alpha_{IIb}\beta_3$, $\alpha_v\beta_3$ and $\alpha_5\beta_1$) (Marcinkiewicz et al., 1997; Wierzbicka-Patinowski et al., 1999). KGD-containing barbourin inhibits the $\alpha_{IIb}\beta_3$ integrin with a high degree of selectivity (Scarborough et al., 1991). The integrin specificity profile of atrolysin E is unknown, although due to its inhibition of ADP-and collagen-stimulated platelet aggregation, $\alpha_{IIb}\beta_3$ may be one of its target receptor(s) (Shimokawa et al., 1998).

Dimeric disintegrins exhibit the largest sequence diversity in their integrin binding motifs. EC3, a heterodimeric disintegrin from *Echis carinatus* (=*E. sochureki*) venom is a selective and potent antagonist of the binding of $\alpha_4\beta_1$ and $\alpha_4\beta_7$ integrins to immobilized VCAM-1 and MAdCAM-1, respectively. It is also a weak inhibitor of $\alpha_5\beta_1$ and $\alpha_{IIb}\beta_3$ integrins and does not inhibit $\alpha_v\beta_3$ integrin (Marcinkiewicz et al., 1999a). The inhibitory activity of EC3 towards the α_4 integrins is associated with the MLD sequence of its B-subunit. The EC3 A-subunit contains a VGD motif, and the ability of EC3 to inhibit $\alpha_5\beta_1$ resides in both subunits. EMF-10, another heterodimeric disintegrin isolated from the venom of *Eristocophis macmahoni*, is an extremely potent and selective inhibitor of integrin $\alpha_5\beta_1$ binding to fibronectin and partially inhibiting the adhesion of cells expressing integrins $\alpha_{IIb}\beta_3$, $\alpha_v\beta_3$ and $\alpha_4\beta_1$ to their appropriate ligands (Marcinkiewicz et al., 1999b). Selective recognition of $\alpha_5\beta_1$ by EMF-10 is associated with the MGD(W) sequence, a motif located in the active loop of the B-subunit, and expression of $\alpha_5\beta_1$ inhibitory activity may also depend on the RGD(N) motif in the A-subunit. The presence of a WGD motif in CC8, a heterodimeric disintegrin isolated from the venom of the North African sand viper, *Cerastes cerastes cerastes*, has been reported to increase its inhibitory effect on $\alpha_{IIb}\beta_3$, $\alpha_v\beta_3$ and $\alpha_5\beta_1$ integrins (Calvete et al., 2002). Fig. 6 displays a diagram of the integrin family and the different tripeptide motifs found in disintegrins that block specific integrin–ligand interactions.

Dimeric disintegrins are widely distributed in *Echis* and *Vipera* venoms, and probably also in the venoms of many other species of *Crotalidae* and *Viperidae*, which are, in addition, rich sources of monomeric disintegrins. It is worth mentioning that non-RGD disintegrins are present in venoms which also contain RGD-disintegrins (Calvete et al., 2003) (Table 1). The large sequence and structural diversity exhibited by the different subfamilies of disintegrins strongly suggests that disintegrins, like toxins from other venoms (Duda and Palumbi, 1999; Kordis et al., 2002; Ohno et al., 2002), have evolved rapidly by adaptative evolution driven by positive Darwinian selection. The accelerated evolution of toxins may be linked to adaptation to the environment, including feeding habits (Okuda et al., 2001). The co-existence in the same snake species of disintegrins with conserved RGD-motif and disintegrins with variable non-RGD sequences support the hypothesis that, following gene duplication, one copy of the gene (i.e. that coding for an RGD disintegrin) divergently evolved under pressure dictated by the ancestral function (blocking

Table 1
Co-existence in snake venoms of RGD- and non-RGD-containing PII disintegrins

Snake species	RGD disintegrin	Non-RGD disintegrin
Echis sochureki carinatus	Echistatin, EC6B, schistatin	EC3A [VGD], EC3B (=EC6B) [MLD]
Echis multisquamatus	Multisquamatin	EMS11 [MLD]
Echis ocellatus	Ocellatusin, EO4A	EO4B (=EO5B) [VGD], EO5A [MLD]
Eristocophis macmahoni	Eristostatin EMF10A	EMF10B [MGD]
Cerastes cerastes cerastes	CC8A	CC8B [WGD]
Vipera lebetina obtusa	VLO4	VLO5A [VGD], VLO5B [MLD], obtustatin [KTS]
Vipera lebetina	Lebein	Lebetase [VGD]
Agkistrodon piscivorus piscivorus	Piscivostatin 2B	Piscivostatin 2A [KGD]
Crotalus atrox	Crotatroxin	Atrolysin E [MVD]

The non-RGD motifs are specified in between square brackets, and their integrin inhibitory specificity is schematized in Fig. 6.

of platelet aggregation). The duplicated gene(s) (non-RGD disintegrin), unencumbered by a functional constraint, is(are) free to search for new physiological roles such as inhibition of non-RGD-dependent integrin receptors.

4. The integrin inhibitory loop: structure-function correlations

NMR studies of short (echistatin, 2ECH, 1RO3; obtustatin, 1MPZ) and medium-sized (kistrin, 1KST; flavoridin, 1FVL; albolabrin, Smith et al., 1996; salmosin, 1IQ2, 1L3X; rhodostomin, 1JYP) disintegrins, and the recent crystal structure of the medium-sized disintegrin trimestatin (1J2L, Fujii et al., 2003) (PDB coordinates available at http://www.rcsb.org/pdb) revealed that a mobile 9–11-residue loop protruding 14–17 Å from the protein core harbours the active tripeptide (Fig. 7). As judged from the consensus sequences of the different disintegrin subfamilies (Fig. 1), the integrin-binding loop represents a mutational 'hot spot' and the only conserved characteristic (except in obtustatin) is the presence of an acidic residue (Asp in short, dimeric, medium-sized, and long disintegrins; Glu or Asp in disintegrin/cysteine-rich domains) at the C-terminal site of the integrin-binding motif. The crystal structure of the extracellular segment of integrin $\alpha_V\beta_3$ in complex with an RGD ligand (Xiong et al., 2002) showed that the peptide fits into a crevice between the α_V propeller and the β_3 A-domain. The Arg side-chain is held in place by interactions with α_V carboxylates 218 and 150, the Gly residue makes several hydrophobic interactions with α_V, and the Asp ligand interacts primarily with βA residues. Thus, the conserved aspartate residue might be responsible for the binding of disintegrins to integrin receptors which share a β subunit, while the two other residues of the integrin-binding motif (RG, MG, WG, ML, VG) may dictate the primary integrin specificity.

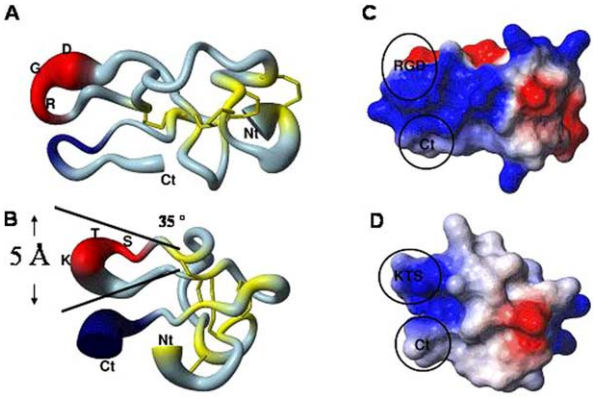


Fig. 7. NMR structures of short disintegrins. Panels **A** and **B** display the solution structures of echistatin (PDB code 1RO3) and obtustatin (PDB code 1MPZ), respectively in the 'sausage' representation where the thickness of the backbone is proportional to the RMSD value of the backbone atoms coordinates. Disulfide bonds and active tri-peptide are colored in yellow and red, respectively. C-terminal tail residues 43–45 are colored in dark blue. The amplitude of lateral movement of the integrin-binding loop (which is similar in both disintegrins) is indicated. Panels **C** and **D** show the surface electrostatic potential of echistatin and obtustatin, respectively. The molecules are in the same orientation as in Panels A and B. The positively and negatively charged areas are colored blue and red, respectively. The integrin recognition motifs (RGD and KTS) and the C-terminal regions of both disintegrins, which form a conformational epitope engaged in extensive interactions with the target integrin receptor, are circled.

Obtustatin, a potent and selective inhibitor of the $\alpha_1\beta_1$ integrin in vitro and of angiogenesis in vivo, isolated from the venom of *Vipera lebetina obtusa* (Marcinkiewicz et al., 2003) is the shortest polypeptide of the disintegrin family and posesses an integrin recognition loop which is two residues shorter than that of other disintegrins and which harbors in a lateral position the $\alpha_1\beta_1$ integrin inhibitory KTS motif (Moreno-Murciano et al., 2003a) (Fig. 7). The NMR solution structure of obtustatin (Moreno-Murciano et al., 2003b) showed that the integrin recognition loop exhibits a global, lateral hinge motion within a range of 35° and with maximum displacement of about 5 Å (Fig. 7). In comparison, the orientational flexibility of the integrin binding loop of echistatin has been reported to occur within a range of 60–70° (Chen et al., 1994). In line with the concept that fast recognition and fitting processes are typically brought about by mobile segments in protein structures (Burgen et al., 1975; Williams, 1989), the loop movement of obtustatin occur in the 100–300 ps timescale and is articulated at residues W20, H27 and Y28 located at the base of the loop. Noteworthy, the side-chain of threonine-22, the middle residue of the KTS motif, which based on synthetic peptides appears to be the most critical residue for expression of the inhibitory activity of obtustatin on the binding of integrin $\alpha_1\beta_1$ to collagen IV (Moreno-Murciano et al., 2003a), is not solvent-exposed but oriented towards the loop center and hydrogen-bonded to residues T25 and S26 (Monleón et al., 2003). This network of interactions restrains the mobility of threonine-22 and suggests that the role of this residue may be linked with a structuring mechanism of the integrin-binding loop for proper recognition of the $\alpha_1\beta_1$ integrin, rather than with a direct role in receptor binding. More structural studies on disintegrins displaying non-Gly residues in the central position of their integrin inhibitory motifs are needed to check the hypothesis that the functional importance of this residue lies in maintaining the active conformation of the integrin-binding loop.

A number of studies support a functional role for the amino acids flanking the integrin-recognition tripeptide in determining the receptor-binding characteristics (Lu et al., 1996; McLane et al., 1996). Disintegrins with RGD(W/F) and RGD(N/D) sequences are quite selective in inhibiting native ligand binding to integrins $\alpha_{IIb}\beta_3$ and $\alpha_v\beta_3$, respectively. Molecular modelling and NMR structure determination of cyclic RGD peptides (Pfaff et al., 1994)

showed the importance of the amino acid immediately C-terminal of the RGD sequence in determining the conformation of the RGD loop. The width and shape of the integrin-binding loop represents an important structural feature that determines integrin binding selectivity, and the distance between the Cβ atoms of Arg and Asp distinctly affects the fitting of disintegrins in the binding pocket of integrins $\alpha_{IIb}\beta_3$ and $\alpha_v\beta_3$, which share the β_3 subunit: the optimum distance is in the range of 7.5–8.5 Å for $\alpha_{IIb}\beta_3$ and at or below 6.7 Å for $\alpha_v\beta_3$ (and $\alpha_5\beta_1$) (Pfaff et al., 1994). Other loop residues located more distantly of the active tripeptide, both downstream (Wierzbicka-Patynowski et al., 1999) and positioned N-terminal to the primary integrin recognition motif, also play pivotal roles in modulating the integrin specificity of disintegrins of both monomeric RGD-bearing (Rahman et al., 1998) and dimeric (Bazan- Socha et al., 2004) MLD-containing disintegrins.

5. Functional epitopes in the C-terminal tail

The NMR solution structures of a number of disintegrins show that their C-terminal portions are in close spatial proximity to the active loop (Fig. 7). In the case of obtustatin, the integrin-binding loop and the C-terminal tail interact through a NOE between K21 H_β and Y39 H_N and displays concerted motions due to a hinge effect articulated at residues W20, H27, Y28, C36, and L38 (Monleón et al., 2003). Similarly, Senn and Klaus (1993) have reported that the C-terminal residues of flavoridin (C64-Trp67) are connected to the RGD-loop. The amino acid residues in the C-terminal region are not highly conserved among disintegrins, further suggesting that this region could act as a secondary determinant of integrin-binding specificity/potency. In this sense, it has been shown that the C-terminal tail of echistatin is involved in modulating the binding affinity of the disintegrin towards integrin $\alpha_{IIb}\beta_3$ triggering the expression of conformational changes (LIBS, ligand-induced binding sites) in the receptor that lead to a further increase of the binding affinity and the inhibitory potency of the disintegrin (Marcinkiewicz et al., 1997). Thus, an echistatin C-terminal peptide inhibited echistatin–integrin $\alpha_{IIb}\beta_3$ binding, activated the integrin to bind immobilized ligands, induced the expression of activation-dependent conformational epitopes, and increased fibrinogen binding to peptide-treated platelets (Wright et al., 1993). In addition, both deletion of the sequence ^{40}PRNP^{43} (Gould et al., 1990) and replacement of echistatin's C-terminal sequence (^{44}HKGPAT^{49}) with that of eristostatin (WNG) decreased the inhibitory potential the mutated echistatin on ADP-induced platelet aggregation (Marcinkiewicz et al., 1997). As a whole, these data indicate that the C-terminal tail may act in synergy with the integrin-binding loop to endow disintegrins with high affinity and selectiveness for integrin receptors. Fujii et al. (2003) have generated a docking model of the disintegrin trimestatin fitted in the RGD-binding site of the structure of the extracellular domains of integrin $\alpha_v\beta_3$ in complex with the cyclic pentapeptide Arg-Gly-Asp-[D-Phe]-[N-methyl-Val] (Xiong et al., 2002). In agreement with biochemical and molecular biology reports, the model revealed that within the RGD-loop, Phe52 adjacent to the ^{49}RGD^{51} sequence form hydrophobic contacts with α_{IIb} Tyr178 and with β_3 Tyr166 and Pro53 has contacts with Asp179 and Arg214 of the β_3 subunit. Furthermore, the side chains of the C-terminal tail residues Arg66, Trp67, Asn68, and the main chain of Asp69, interact with amino acids mainly of the α_{IIb} propeller domain (Thr116, Lys119, Glu121, Asp148) but also with residues Tyr166 and Asp179 of the β_3 A-domain (see Fig. 3 in Fujii et al., 2003).

Recently, we have revised the structure and dynamics of echistatin using homonuclear NMR methods

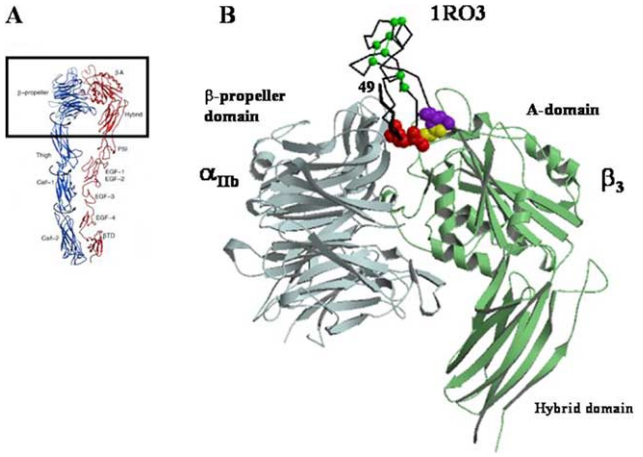


Fig. 8. Echistatin-integrin $\alpha_{IIb}\beta_3$ interaction. A. Model of the extracellular region of $\alpha_v\beta_3$ showing its multidomain molecular architecture (Xiong et al., 2001). Representation of echistatin docked onto a model of the globular head of integrin $\alpha_{IIb}\beta_3$ based on the crystal structure of of $\alpha_v\beta_3$ (PDB code 1L5G) (Feuston et al., 2003). The coordinates of the $\alpha_{IIb}\beta_3$ model were kindly provided by Dr Feuston. The amino acids of the RGD motif are shown in the space-filling model in magenta, yellow and red, respectively. Disulfide bonds are depicted in green, and the C-terminal tail is drawn as a thick line. The C-terminal residue of echistatin (Thr49) is labeled.

(Monleón, Esteve, Kovacs, Calvete, Celda, submitted). In the 100–300 ps time-scale, the integrin binding loop of echistatin (PDB accession code 1RO3) displays an overall movement with maximal amplitude of 30°, which is articulated at residues Gly17, Thr18, and Asp29 located at the base of the loop. Superimposed to this loop motion, the backbone atoms of the active tripeptide amino acids show local enhanced flexibility due to a hinge effect at flanking residues Ala23 and Asp27. The full-length C-terminal polypeptide is visible and folds as a β-hairpin running parallel to the RGD loop and exposing at the tip residues Pro43, His44, and Lys45. A network of long-range NOEs indicates the existence of concerted motions between the C-terminal tail and the RGD loop. An integrin-disintegrin complex developed using a model for the extracellular domains of $\alpha_{IIb}\beta_3$ (Feuston et al., 2003) (Fig. 8) and the refined solution structure of echistatin (1RO3) (Fig. 7) clearly shows that, in agreement with previous biochemical and mutational data, the RGD loop and the C-terminal region of echistatin form a conformational epitope engaged in extensive interactions with the target integrin receptor, and provide the molecular basis for understanding the functional synergy between these two functional epitopes.

6. Concluding remarks

Disintegrins are small, cysteine-rich integrin inhibitors, which evolved rapidly by adaptative evolution from a cell membrane ADAM ancestor in the venom gland of vipers and rattlesnakes after the separation of reptiles and birds. The structural and functional complexity of disintegrins contrasts with their small molecular size. Though a tripeptide at the tip of a mobile loop represents the primary integrin inhibitory determinant, the potency and receptor selectivity of disintegrins is finely tuned by residues flanking the active sequence. We hypothesize that the central residue of the primary integrin-binding motif is important for maintaining the conformation of the active loop. The C-terminal region of disintegrins displays concerted motions with the integrin recognition loop, and harbors additional conformational epitopes engaged in extensive interactions with the receptor. The composition (surface potential), architecture, and dynamics of the integrin-interacting disintegrin patch, formed by the active loop and C-terminal tail epitopes, may determine the selective inhibition of integrin receptors.

Acknowledgements

We wish to thank Dr Bradley P. Feuston (Merck Research Laboratories, West Point, PA, USA) for the PDB coordinates of the $\alpha_{IIb}\beta_3$ model. Current research on structure-function correlations of disintegrins carried out in the authors laboratories has been financed by grant BMC2001-3337 from the Ministerio de Ciencia y Tecnología, Madrid, Spain. PJ and LS are recipients of a pre-doctoral fellowship (FPI, formación de personal investigador) from the Spanish Ministerio de Ciencia y Tecnología, and an I3P contract, respectively.

References

Bazan-Socha, S., Kisiel, D.G., Young, B., Theakston, R.D.G., Calvete, J.J., Sheppard, D., Marcinkiewicz, C., 2004. Structural requirements of MLD-containing disintegrins for functional interactions with $\alpha_4\beta_1$ and $\alpha_9\beta_1$ integrins. Biochemistry 43, 1639–1647.

Bilgrami, S., Tomar, S., Yadav, S., Kaur, P., Kumar, J., Jabeen, T., Sharma, S., Sinhg, T.P., 2004. Crystal structure of schistatin, a disintegrin homodimer from saw-scaled viper (*Echis carinatus*) at 2.5 Å resolution. J. Mol. Biol. 341, 829–837.

Bjarnason, J.B., Fox, J.W., 1994. Hemorrhagic metalloproteinases from snake venoms. Pharmac. Ther. 62, 325–372.

Burgen, A.S.V., Roberts, G.C.K., Freeney, J., 1975. Binding of flexible ligands to macromolecules. Nature 253, 753–755.

Calvete, J.J., Schäfer, W., Soszka, T., Lu, W., Cook, J.J., Jameson, B.A., Niewiarowski, S., 1991. Identification of the disulfide bond pattern in albolabrin, an RGD-containing peptide from the venom of *Trimeresurus albolabris*: significance for the expression of platelet aggregation inhibitory activity. Biochemistry 30, 5225–5229.

Calvete, J.J., Moreno-Murciano, M.P., Sanz, L., Jürgens, M., Schrader, M., Raida, M., Benjamin, D.C., Fox, J.W., 2000a. The disulfide bond pattern of catrocollastatin C, a disintegrin/cysteine-rich protein isolated from *Crotalus atrox* venom. Protein Sci. 9, 1365–1373.

Calvete, J.J., Jürgens, M., Marcinkiewicz, C., Romero, A., Schrader, M., Niewiarowski, S., 2000b. Disulfide bond pattern and molecular modelling of the dimeric disintegrin EMF-10, a potent and selective integrin $\alpha_5\beta_1$ antagonist from *Eristocophis macmahoni* venom. Biochem. J. 345, 573–581.

Calvete, J.J., Fox, J.W., Agelan, A., Niewiarowski, S., Marcinkiewicz, C., 2002. The presence of the WGD motif in CC8 heterodimeric disintegrin increases its inhibitory effect on $\alpha_{IIb}\beta_3$, $\alpha_v\beta_3$, and $\alpha_5\beta_1$ integrins. Biochemistry 41, 2014–2021.

Calvete, J.J., Moreno-Murciano, M.P., Theakston, R.D.G., Kisiel, D.G., Marcinkiewicz, C., 2003. Snake venom disintegrins: novel dimeric disintegrins and structural diversification by disulfide bond engineering. Biochem. J. 372, 725–734.

Chen, Y., Suri, A.K., Kominos, D., Sanyal, G., Naytor, A.M., Pitzenberger, S.M., Garsky, V.M., Levy, R.M., Baum, J., 1994. Three-dimensional structure of echistatin and dynamics of the active site. J. Biomol. NMR 4, 307–324.

Duda jr., T.F., Palumbi, S.R., 1999. Developmental shifts and species selection in gastropods. Proc. Natl Acad. Sci. USA 96, 6820–6823.

Feuston, B.P., Culberson, J.C., Hartman, G.D., 2003. Molecular model of the $\alpha_{IIb}\beta_3$ integrin. J. Med. Chem. 46, 5316–5325.

Fox, J.W., Long, C., 1998. The ADAMs/MDC family of proteins and their relationships to the snake venom metalloproteinases. In: Bailey, G.S. (Ed.), Snake Venom Enzymes. Alaken Press, Ft. Collins, Co, USA, pp. 151–178.

Fry, B.G., 1999. Structure-function properties of venom components from Australian elapids. Toxicon 37, 11–32.

Fujii, Y., Okuda, D., Fujimoto, Z., Horii, K., Morita, T., Mizuno, H., 2003. Crystal structure of trimestatin, a disintegrin containing a cell adhesion recognition motif RGD. J. Mol. Biol. 332, 1115–1122.

Gould, R.J., Polokoff, M.A., Friedman, P.A., Huang, T-F., Holt, J.C., Cook, J.J., Niewiarowski, S., 1990. Disintegrins: a family of integrin inhibitory proteins from viper venoms. Proc. Soc. Exp. Biol. Med. 195, 168–171.

Hite, L.A., Shannon, J.D., Bjarnason, J.B., Fox, J.W., 1992. Sequence of a cDNA clone encoding the zinc metalloproteinase hemorrhagic toxin e from *Crotalus atrox*: evidence for signal, zymogen, and disintegrin-like structures. Biochemistry 31, 6203–6211.

Hite, L.A., Jia, L-G., Bjarnason, J.B., Fox, J.W., 1994. cDNA sequences of four snake venom metalloproteinases: structure, classification, and their relationship to mammalian reproductive proteins. Arch. Biochem. Biophys. 308, 182–191.

Hong, S.-Y., Sohn, Y.-D., Chung, K.-H., Kim, D.-S., 2002. Structural and functional significance of disulfide bonds in saxatilin, a 7.7 kDa disintegrin. Biochem. Biophys. Res. Commun. 293, 530–536.

Huang, T.-F., Holt, J.C., Lukasiewicz, H., Niewiarowski, S., 1987. Trigramin. A low molecular weight peptide inhibiting fibrinogen interaction with platelet receptors expressed on glycoprotein IIb–IIIa complex. J. Biol. Chem. 262, 16157–16163.

Hynes, R.O., 2002. Integrins: bidirectional, allosteric signaling machines. Cell 110, 673–687.

Ivaska, J., Kapyla, J., Pentikainen, O., Hoffrén, A.M., Hermonen, J., Huttunen, P., Johnson, M.S., Heino, J., 1999. A peptide inhibiting the collagen binding function of integrin α_2I domain. J. Biol. Chem. 274, 3513–3521.

Jia, L-G., Shimokawa, K-i, Bjarnason, J.B., Fox, J.W., 1996. Snake venom metalloproteinases: structure, function and relationship to the ADAMs family of proteins. Toxicon 34, 1269–1276.

Jia, L.-G., Wang, X.-M., Shannon, J.D., Bjarnason, J.B., Fox, J.W., 1997. Function of disintegrin-like/cysteine-rich domains of atrolysin A. Inhibition of platelet aggregation by recombinant protein and peptide antagonists. J. Biol. Chem. 272, 13094–13102.

Jia, L.-G., Wang, X.-M., Shannon, J.D., Bjarnason, J.B., Fox, J.W., 2000. Inhibition of platelet aggregation by the recombinant cysteine-rich domain of the hemorrhagic snake venom metalloproteinase, atrolysin A. Arch. Biochem. Biophys. 373, 281–286.

Juárez, P., Sanz, L., Calvete, J.J., 2004. Snake venomics: characterization of protein families in *Sistrurus barbouri* venom by cysteine mapping, N-terminal sequencing, and MS/MS analysis. Proteomics 4, 327–338.

Kamiguti, A.S., Hay, C.R.M., Zuzel, M., 1995. Inhibition of collage-induced platelet aggregation as the result of cleavage of $\alpha_2\beta_1$ integrin by the snake venom metalloprotease jararhagin. Biochem. J. 320, 635–641.

Kamiguti, A.S., Gallagher, P., Marcinkiewicz, C., Theakston, R.D.G., Zuzel, M., Fox, J.W., 2003. Identification of sites in the cysteine-rich domain of the class P-III snake venom metalloproteinases responsible for inhibition of platelet function. FEBS Lett. 549, 129–134.

Kini, R.M., Evans, H.J., 1992. Structural domains in venom proteins: evidence that metalloproteinases and nonenzymatic platelet aggregation inhibitors (disintegrins) from snake venoms are derived by proteolysis from a common precursor. Toxicon 30, 265–293.

Kordis, D., Krizaj, I., Gubensek, F., 2002. Functional diversification of animal toxins by adaptative evolution. In: Ménez, A. (Ed.), Perspectives in Molecular Toxinology. Wiley, New York, pp. 401–419.

Lu, X., Rahman, S., Kakkar, V.V., Authi, K.S., 1996. Substitutions of proline 42 to alanine and methionine 46 to asparagine around the RGD domain of the neurotoxin dendroaspin alter its preferential antagonism to that resembling the disintegrin elegantin. J. Biol. Chem. 271, 289–294.

Marcinkiewicz, C., Vijay-Kumar, S., McLane, M.A., Niewiarowski, S., 1997. Significance of the RGD loop and C-terminal domain of echistatin for recognition of $\alpha_{IIb}\beta_3$ and $\alpha_v\beta_3$ integrins and expression of ligand-induced binding sites. Blood 90, 1565–1575.

Marcinkiewicz, C., Calvete, J.J., Marcinkiewicz, M.M., Raida, M., Vijay-Kumar, S., Huang, Z., Lobb, R.R., Niewiarowski, S., 1999a. EC3, a novel heterodimeric disintegrin from *Echis carinatus* venom, inhibits α_4 and α_5 integrins in an RGD-independent manner. J. Biol. Chem. 274, 12468–12473.

Marcinkiewicz, C., Calvete, J.J., Vijay-Kumar, S., Marcinkiewicz, M.M., Raida, M., Schick, P., Lobb, R.R., Niewiarowski, S., 1999b. Structural and functional characterization of EMF10, a heterodimeric disintegrin from *Eristocophis macmahoni* venom that selectively inhibits $\alpha_5\beta_1$ integrin. Biochemistry 38, 13302–13309.

Marcinkiewicz, C., Taooka, Y., Yokosaki, Y., Calvete, J.J., Marcinkiewicz, M.M., Lobb, R.R., Niewiarowski, S., Sheppard, D., 2000. Inhibitory effects of MLDG-containing heterodimeric disintegrins reveal distinct structural requirements for interaction of the integrin $\alpha_9\beta_1$ with VCAM-1, tenascin-C, and osteopontin. J. Biol. Chem. 275, 31930–31937.

Marcinkiewicz, C., Weinreb, P.H., Calvete, J.J., Kisiel, D.G., Mousa, S.A., Tuszynski, G.P., Lobb, R.R., 2003. Obtustatin: a potent selective inhibitor of $\alpha_1\beta_1$ integrin *in vitro* and angiogenesis *in vivo*. Cancer Res. 63, 2020–2023.

Markland, F.S., 1998. Snake venoms and the hemostatic system. Toxicon 36, 1749–1800.

McLane, M.A., Vijay-Kumar, S., Marcinkiewicz, C., Calvete, J.J., Niewiarowski, S., 1996. Importance of the structure of the RGD-containing loop in the disintegrins echistatin and eristostatin for recognition of $\alpha_{IIb}\beta_3$ and $\alpha_v\beta_3$ integrins. FEBS Lett. 291, 139–143.

McLane, M.A., Marcinkiewicz, C., Vijay-Kumar, S., Wierzbicka-Patynowski, I., Niewiarowski, S., 1998. Viper venom disintegrins and related molecules. Proc. Soc. Exp. Biol. Med. 219, 109–119.

Ménez, A. (Ed.), 2002. Perspectives in Molecular Toxinology. Wiley, Chichester, UK.

Monleón, D., Moreno-Murciano, M.P., Kovacs, H., Marcinkiewicz, C., Calvete, J.J., Celda, B., 2003. Concerted motions of the integrin-binding loop and the c-terminal tail of the non-RGD disintegrin obtustatin. J. Biol. Chem. 278, 45570–45576.

Moreno-Murciano, M.P., Monleón, D., Calvete, J.J., Celda, B., Marcinkiewicz, C., 2003a. Amino acid sequence and homology modeling of obtustatin, a novel non-RGD-containing short disintegrin isolated from the venom of *Vipera lebetina obtusa*. Protein Sci. 12, 366–371.

Moreno-Murciano, M.P., Monleón, D., Marcinkiewicz, C., Calvete, J.J., Celda, B., 2003b. NMR solution structure of the non-RGD disintegrin obtustatin. J. Mol. Biol. 329, 135–145.

Moura da Silva, A.M., Theakston, R.D.G., Crampton, J.M., 1996. Evolution of disintegrin cysteine-rich and mammalian matrix-degrading metalloproteinases: gene duplication and divergence of a common ancestor rather than convergent evolution. J. Mol. Evol. 43, 263–269.

Moura da Silva, A.M., Marcinkiewicz, C., Marcinkiewicz, M., Niewiarowski, S., 2001. Selective recognition of $\alpha_2\beta_1$ integrin by jararhagin, a metalloproteinase/disintegrin from *Bothrops jararaca* venom. Thromb. Res. 102, 153–159.

Niewiarowski, S., Marcinkiewicz, C., Wierzbicka-Patynowski, I., McLane, M.A., Calvete, J.J., 2002. Structure and function of disintegrins and C-lectins: viper venom proteins modulating cell adhesion. In: Ménez, A. (Ed.), Perspectives in Molecular Toxinology. Wiley, New York, pp. 327–340.

Nikai, T., Taniguchi, K., Komori, Y., Masuda, K., Fox, J.W., Sugihara, H., 2000. Primary structure and functional characterization of bilitoxin-1, a novel dimeric P-II snake venom metalloproteinase from *Agkistrodon bilineatus* venom. Arch. Biochem. Biophys. 378, 6–15.

Nymalm, Y., Puranen, J.S., Nyholm, T.K.M., Käpylä, J., Kidron, H., Pentikäinen, O.T., Airenne, T.T., Heino, J., Slotte, J.P., Johnson, M.S., Salminen, T.A., 2004. Jararhagin-derived RKKH peptides induce structural changes in α_1?I domain of human integrin $\alpha_1\beta_1$. J. Biol. Chem. 279, 7962–7970.

Ohno, M., Ogawa, T., Oda-Ueda, N., Chijiwa, T., Hattori, S., 2002. Accelerated and regional evolution of snake venom gland isozymes. In: Ménez, A. (Ed.), Perspectives in Molecular Toxinology. Wiley, New York, pp. 387–419.

Okuda, D., Nozaki, C., Sekiya, F., Morita, T., 2001. Comparative biochemistry of disintegrins isolated from snake venoms: consideration of the taxonomy and geographical distribution of snakes in the genus *Echis*. J. Biochem. 129, 615–620.

Okuda, D., Koike, H., Morita, T., 2002. A new gene structure of the disintegrin family: a subunit of dimeric disintegrin has a short coding region. Biochemistry 41, 14248–14254.

Oshikawa, K., Terada, S., 1999. Ussuristatin 2, a novel KGD-bearing disintegrin from *Agkistrodon ussuriensis* venom. J. Biochem. 125, 31–35.

Pfaff, M., Tangemann, K., Müller, B., Gurrath, M., Müller, G., Kessler, H., Timpl, R., Engel, J., 1994. Selective recognition of cyclic RGD peptides of NMR defined conformation by $\alpha_{IIb}\beta_3$ and $\alpha_v\beta_3$, and $\alpha_5\beta_1$ integrins. J. Biol. Chem. 269, 20233–20238.

Rahman, S., Aitken, A., Flynn, G., Formstone, C., Savidge, G.F., 1998. Modulation of RGD sequence motifs regulates disintegrin recognition of $\alpha_{IIb}\beta_3$ and $\alpha_5\beta_1$ integrin complexes. Biochem. J. 335, 247–257.

Scarborough, R.M., Rose, J.W., Hsu, M.A., Phillips, D.R., Fried, V.A., Campbell, A.M., Nannizzi, L., Charo, I.F., 1991. Barbourin. A GPIIb-IIIa-specific integrin antagonist from the venom of *Sistrurus M. barbouri*. J. Biol. Chem. 266, 9359–9362.

Senn, H., Klaus, W., 1993. The nuclear magnetic resonance solution structure of flavoridin, an antagonist of the platelet GPIIb-IIIa receptor. J. Mol. Biol. 232, 907–925.

Shimokawa, K-I., Shannon, J.D., Jia, L-G., Fox, J.W., 1997. Sequence and biological activity of catrocollastatin-C: a disintegrin-like/cysteine-rich two domain protein from *Crotalus atrox* venom. Arch. Biochem. Biophys. 343, 35–43.

Shimokawa, K-i., Jia, L-G., Shannon, J.D., Fox, J.W., 1998. Isolation, sequence analysis, and biological activity of atrolysin E/D, the non-RGD disintegrin domain from *Crotalus atrox* venom. Arch. Biochem. Biophys. 354, 239–246.

Shin, J., Hong, S-Y., Chung, K., Kang, I., Jang, Y., Kim, D.-S., Lee, W., 2003. Solution structure of a novel disintegrin, salmosin, from *Agkistrodon halys* venom. Biochemistry 42, 14408–14415.

Smith, K.J., Jaseja, M., Lu, X., Williams, J.A., Hyde, E.I., Trayer, I.P., 1996. Three-dimensional structure of the RGD-containing snake toxin albolabrin in solution, based on ^{1}H NMR spectroscopy and simulated annealing calculations. Int. J. Pept. Protein Res. 48, 220–228.

Souza, D.H.F., Iemma, M.R.C., Ferreira, L.L., Faria, J.P., Oliva, M.L.V., Zingali, R.B., Niewiarowski, S., Selistre de Araujo, H.S., 2000. The disintegrin-like domain of the snake venom metalloprotease alternagin inhibits $\alpha_2\beta_1$ integrin-mediated cell adhesion. Arch. Biochem. Biophys. 384, 341–350.

Wierzbicka-Patynowski, I., Niewiaroeski, S., Marcinkiewicz, C., Calvete, J.J., Marcinkiewicz, M.M., McLane, M.A., 1999. Structural requirements of echistatin for the recognition of $\alpha_v\beta_3$ and $\alpha_5\beta_1$ integrins. J. Biol. Chem. 274, 37809–37814.

Williams, R.J.P., 1989. NMR studies of mobility within protein structure. Eur. J. Biochem. 183, 479–497.

Wright, P.S., Saudek, V., Owen, T.J., Harbeson, S.L., Bitonti, A.J., 1993. An echistatin C-terminal peptide activates GPIIbIIIa binding to fibrinogen, fibronectin, vitronectin and collagen type I and type IV. Biochem. J. 293, 263–267.

Wu, W.B., Chang, S.C., Liau, M.Y., Huang, T.F., 2001. Purification, molecular cloning and mechanism of action of graminelysin I, a snake-venom-derived metalloproteinase that induces apoptosis of human endothelial cells. Biochem. J. 357, 719–728.

Xiong, J.-P., Stehle, T., Diefenbach, B., Zhang, R., Dunker, R., Scott, D.L., Joachimiak, A., Goodman, S.L., Arnaout, M.A., 2001. Crystal structure of the extracellular segment of integrin $\alpha_v\beta_3$. Science 294, 339–345.

Xiong, J.-P., Stehle, T., Zhang, R., Joachimiak, A., Frech, M., Goodman, S.L., Arnaout, M.A., 2002. The crystal structure of the extracellular segment of integrin $\alpha_v\beta_3$ in complex with an Arg-Gly-Asp ligand. Science 296, 151–155.

Zhou, Q., Smith, J.B., Grossman, M.H., 1995. Molecular cloning and expression of catrocollastatin, a snake-venom protein from *Crotalus atrox* (western diamondback rattlesnake) which inhibits platelet adhesion to collagen. Biochem. J. 307, 411–417.

Zigrino, P., Kamiguti, A.S., Eble, J., Drescher, C., Nischt, R., Fox, J.W., Mauch, C., 2002. The reprolysin jararhagin, a snake venom metalloproteinase, functions as a fibrillar collagen agonist involved in fibroblast cell adhesion and signaling. J. Biol. Chem. 277, 40528–40535.

5.2. TRABAJO 2: cDNA cloning and functional expression of Jerdostatin, a novel RTS-disintegrin from *Trimeresurus jerdonii* and a specific antagonist of the $\alpha_1\beta_1$ integrin

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 280, NO. 49, pp. 40714–40722, December 9, 2005

cDNA Cloning and Functional Expression of Jerdostatin, a Novel RTS-disintegrin from *Trimeresurus jerdonii* and a Specific Antagonist of the $\alpha_1\beta_1$ Integrin*

Received for publication, September 6, 2005 Published, JBC Papers in Press, October 7, 2005, DOI 10.1074/jbc.M509738200

Libia Sanz[‡1], Run-Qiang Chen[§1], Alicia Pérez[‡], Rebeca Hilario[‡], Paula Juárez[‡], Cezary Marcinkiewicz[¶], Daniel Monleón[‖], Bernardo Celda[‖], Yu-Liang Xiong[§], Enrique Pérez-Payá[**], and Juan J. Calvete[‡2]

*From the [‡]Instituto de Biomedicina de Valencia, C.S.I.C., Jaime Roig 11, 46010 Valencia, Spain, the [§]Department of Animal Toxinology, Kumming Institute of Zoology, The Chinese Academy of Sciences, Kumming 650223, Peoples Republic of China, the [¶]College of Science and Technology, Center for Neurovirology and Cancer Biology, Temple University, Philadelphia, Pennsylvania 19122, the [‖]Departamento de Química Física, Universitat de València, Dr. Moliner 50, 46100 Valencia, Spain, and the **Centro de Investigación "Principe Felipe" and Consejo Superior de Investigaciones Científicas, Avda. del Saler 16, 46013 Valencia, Spain*

Jerdostatin represents a novel RTS-containing short disintegrin cloned by reverse transcriptase-PCR from the venom gland mRNA of the Chinese Jerdons pit viper *Trimeresurus jerdonii*. The jerdostatins precursor cDNA contained a 333-bp open reading frame encoding a signal peptide, a pre-peptide, and a 43-amino acid disintegrin domain, whose amino acid sequence displayed 80% identity with that of the KTS-disintegrins obtustatin and viperistatin. The jerdostatin cDNA structure represents the first complete open reading frame of a short disintegrin and points to the emergence of jerdostatin from a short-coding gene. The different residues between jerdostatin and obtustatin/viperistatin are segregated within the integrin-recognition loop and the C-terminal tail. Native jerdostatin (r-jerdostatin-R21) and a R21K mutant (r-jerdostatin-K21) were produced in *Escherichia coli*. In each case, two conformers were isolated. One-dimensional ^{1}H NMR showed that conformers 1 and 2 of r-jerdostatin-R21 represent, respectively, well folded and unfolded proteins. The two conformers of the wild-type and the R21K mutant inhibited the adhesion of α_1-K562 cells to collagen IV with IC_{50} values of 180 and 703 nM, respectively. The IC_{50} values of conformers 2 of r-jerdostatin-R21 and r-jerdostatin-K21 were, respectively, 5.95 and 12.5 μM. Neither r-jerdostatin-R21 nor r-jerdostatin-K21 showed inhibitory activity toward other integrins, including $\alpha_{IIb}\beta_3$, $\alpha_v\beta_3$, $\alpha_2\beta_1$, $\alpha_5\beta_1$, $\alpha_4\beta_1$, $\alpha_6\beta_1$, and $\alpha_9\beta_1$ up to a concentration of 24 μM. Although the RTS motif appears to be more potent than KTS inhibiting the $\alpha_1\beta_1$ integrin, r-jerdostatin-R21 is less active than the KTS-disintegrins, strongly suggesting that substitutions outside the integrin-binding motif and/or C-terminal proteolytic processing are responsible for the decreased inhibitory activity.

The integrin family of cell adhesion proteins promotes the attachment and migration of cells on the surrounding extracellular matrix (1, 2). Through signals transduced upon integrin ligation by extracellular matrix proteins, several integrins play key roles in promoting angiogenesis and tumor metastasis (3). However, although antagonists of several integrins (*e.g.* $\alpha_5\beta_1$, $\alpha_v\beta_3$, and $\alpha_v\beta_5$, the primary targets of endostatin, an endogenous negative regulator of angiogenesis (4)) are now under evaluation in clinical trials to determine their potential as therapeutics for cancer and other diseases (5, 6), the precise regulation and exact action of integrins is still unclear (7, 8). Thus, the integrins $\alpha_1\beta_1$ and $\alpha_2\beta_1$ are highly up-regulated by vascular endothelial growth factor in cultured endothelial cells, resulting in an enhanced $\alpha_1\beta_1$- and $\alpha_2\beta_1$-dependent cell spreading on collagen and it has been reported that these integrins provide critical support for vascular endothelial growth factor signaling, endothelial cell migration, and tumor angiogenesis (9). The $\alpha_1\beta_1$ and $\alpha_2\beta_1$ integrins are highly expressed on the microvascular endothelial cells, and blocking of their adhesive properties by monoclonal antibodies (9, 10) or by the snake venom disintegrin obtustatin (11) significantly reduced the vascular endothelial growth factor-driven neovascularization ratio and tumor growth in animal models. Moreover, null-mice lacking integrin $\alpha_1\beta_1$ develop normally, but exhibit reduced vascularity of the skin (9) and have reduced number and size of intratumoral capillaries (12). Ongoing studies with α_2 knock-out mice also suggest a critical role in angiogenesis for the $\alpha_2\beta_1$ integrins (13, 14). Thus, inhibitors of the $\alpha_1\beta_1$ and $\alpha_2\beta_1$ integrins alone or in combination with antagonists of other integrins involved in angiogenesis may prove beneficial in the control of tumor neovascularization.

$\alpha_1\beta_1$ and $\alpha_2\beta_1$ belong to the I-domain bearing subfamily of integrins, and specifically interact with collagen (15). However, despite sharing large structural homology, these two integrins have distinct collagen binding preferences: $\alpha_1\beta_1$ integrin is a very selective receptor of basement membrane type IV collagen, whereas $\alpha_2\beta_1$ is highly specific for fibrillar collagen types I-III (16, 17). Substitution of the cytoplasmic domains of the α_1 and α_2 subunits in transfected human mammary epithelial cells revealed that the two integrins participate in different signal transduction pathways (18). Noteworthy, the $\alpha_1\beta_1$ and $\alpha_2\beta_1$ integrins are the targets of snake venom toxins belonging to different protein families. C-type lectin-like proteins include selective and potent (*i.e.* EMS16 from *Echis multisquamatus*; IC_{50} = 6 nM) inhibitors of $\alpha_2\beta_1$ (19, 20), whereas the only to date known snake venom proteins that specifically antagonize the function of the $\alpha_1\beta_1$ integrin are the disintegrins obtustatin (IC_{50} = 2 nM) from the venom of *Vipera lebetina obtusa* (11, 21), viperistatin (IC_{50} = 0.08 nM) from *Vipera palestinae* (22) and lebestatin (IC_{50} = 0.4 nM) from *Macrovipera lebetina*.[3]

The crystal structure of EMS16 in complex with the integrin α_2 I-do-

* This work was supported in part by National Institutes of Health Grant RO1 CA100145-01A1 and American Heart Association Grant 0230163N (to C. M.), and Ministerio de Educación y Ciencia, Madrid, Spain, Grant BFU2004–01432/BMC (to J. J. C.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[1] Both authors contributed equally to this work.

[2] To whom correspondence should be addressed. Tel.: 34-96-339-1778; Fax: 34-96-369-0800; E-mail: jcalvete@ibv.csic.es.

[3] M. El Ayeb, personal communication.

main has provided insight into the structural basis of the integrin inhibitory specificity of this C-type lectin protein (23). On the other hand, the primary $\alpha_1\beta_1$ binding motif of obtustatin and viperistatin is a KTS tripeptide located in a lateral position of the mobile disintegrins active loop (24), which displays concerted motions with the C-terminal region (25). Now, we report the molecular cloning, primary structure, recombinant expression, and integrin inhibitory characteristics of two conformers of jerdostatin, an RTS-containing disintegrin from the Chinese Jerdon pit viper *Trimeresurus jerdonii*, and of a R21K mutant. In cell adhesion assays, the recombinant (r-) jerdostatin[4] conformers of both the wild-type (r-jerdostatin-R21) and the R21K mutant (r-jerdostatin-K21) selectively blocked, albeit with different potency, the adhesion of K562 cells expressing the integrin $\alpha_1\beta_1$ to collagen IV.

EXPERIMENTAL PROCEDURES

Materials—Human vitronectin was purchased from Chemicon (Temecula, CA). Purified, human collagen IV was provided by Dr. A. Fertala (Thomas Jefferson University, Philadelphia, PA). Highly purified human fibrinogen was a gift from Dr. A. Budzynski (Temple University, Philadelphia, PA). Recombinant human VCAM-1/Ig was a generous gift of Dr. Roy R. Lobb (Biogen). Human fibronectin and laminin were purchased from Sigma. ExTaq® DNA polymerase, dNTP, and DNA marker were from TaKaRa Biotechnology Co., Ltd. (Dalian). PolyATtract® System 1000 kit and the Reverse Transcription System kit were from Promega Biotech.

Cell Lines—K562 cells transfected with α_1, α_2, and α_6 integrins were provided by Dr. M. Hemler (Dana Farber Cancer Institute, Boston, MA). JY cells expressing $\alpha_v\beta_3$ were a gift from Dr. Burakoff (Dana Farber Cancer Institute, Boston, MA). α9- and mock-transfected SW480 cells were generated as described (26). K562 and Jurkat cell lines, which express $\alpha_5\beta_1$ and $\alpha_4\beta_1$ integrins, respectively, were purchased from ATCC (Manassas, VA).

PCR Amplification of Jerdostatin cDNA—The *T. jerdonii* venom glands were collected from Yiliang, Yunnan, China. Isolation of mRNA and reverse transcription was conducted using the PolyATtract System 1000 kit and Reverse Transcription System kit, respectively, according to the manufacturer's protocols. DNA was amplified by PCR using total reverse transcriptase-PCR products as template. The forward primer, 5′-CCAAATCCAG(C/T)CTCCAAAATG-3′, and the reverse primer, 5′-TTCCA(G/T)CTCCATTGTTG(G/T)TTA-3′, were designed according to the highly conserved 5′- and 3′-noncoding regions of the cDNAs encoding for elegantin-2a from *Trimeresurus elegans* (GenBank® accession number AB059572), elegantin-1a from *T. elegans* (GenBank accession number AB059571), and HR2a from *Trimeresurus flavoviridis* (27). The PCR amplification protocol included 35 cycles of 94 °C for 30 s, 50 °C for 30 s, and 72 °C for 2 min. The recovered PCR products were cloned into PMD18-T vector (TaKaRa), and then transformed into *Escherichia coli* strain JM109. The white transformants were screened by PCR and the positive clones were subjected to sequencing on an Applied Biosystems model 377 DNA sequencing system.

Generation of r-Jerdostatin R21K Mutant—Site-directed mutagenesis was performed essentially as described in the QuikChange® site-directed mutagenesis kit of Stratagene (La Jolla, CA). To this end, plasmid pET-32a containing the wild-type jerdostatin sequence flanked by NcoI and XhoI restriction sites was used as the template in the PCR (denaturation at 94 °C for 2 min, followed by 12 cycles of denaturation (30 s at 94 °C), annealing (60 s at 55 °C), and extension (12 min at 68 °C), and a final extension for 10 min at 68 °C) using the forward primer 5′-GGAACAACATGCTGG**AAA**ACCAGTGTATCAAGTCATTAC-TGC-3′ and the reverse primer 5′-ACTTGATACACTGGT**TTT**C-CAGCATGTTGTTCCTGCCGGC-3′ in which the Arg codon AGA has been substituted AAA (Lys) (in boldface). The mutant DNA was sequenced to confirm the absence of undesired mutations.

Synthetic Peptides—Individual peptides and a library of peptides representing the entire integrin-recognition loop of obtustatin but differing in the residue at a single position ($^{19}CX_1$KTSLTSHYC^{29}; CWX_2TSLTSHYC; etc., where X_n is an equimolar mixture of all amino acids except cysteine) were prepared by manual simultaneous multiple peptide synthesis using standard *N*-(9-fluorenyl)methoxycarbonyl (Fmoc) chemistry as described (28). The mixtures at positions "X" were incorporated by coupling a mixture of 19 L-amino acids (cysteine was omitted), with the relative ratio suitability adjusted to yield close to equimolar incorporation. The quality of the synthesized peptide mixtures was validated by mass spectrometry. Individual peptides were purified by preparative reverse phase-HPLC. Peptide identity was confirmed by matrix-assisted laser-desorption ionization time-of-flight (MALDI-TOF) mass spectrometry. The following molar absorption coefficients (ϵ) at 280 nm were used for quantification of peptide mixtures (29): ϵ for the control peptide CWKTSLTSHYC = 5600 (Trp) + 1500 (Tyr) = 7100 M^{-1} cm^{-1}; ϵ for the peptide library CX_2KTSLTSHYC = 1500 + (1/19 × 5600) + (1/19 × 1500) = 1873.7 M^{-1} cm^{-1}; for peptide libraries with X at positions 2–8, ϵ = 5600 + 1500 + (1/19 × 5600) + (1/19 × 1500) = 7473.7 M^{-1} cm^{-1}; and ϵ for the peptide library CWKTSLTSHX_9C = 5600 + (1/19 × 5600) + (1/19 × 1500) = 5973.7 M^{-1} cm^{-1}.

Purification of KTS-disintegrins—Obtustatin and viperistatin were purified from the venoms of *V. lebetina obtusa* and *V. palestinae*, respectively, using the previously described two-step reversed-phase HPLC (11, 21, 22). The purity of the disintegrins was assessed by SDS-PAGE. The monoisotopic masses of the purified disintegrins were determined either by electrospray ionization mass spectrometry with a triple quadrupole-ion trap hybrid instrument (QTrap from Applied Biosystems) equipped with a nanospray source (Protana, Denmark) or by MALDI-TOF mass spectrometry (MS) using an Applied Biosystems DE-Pro spectrometer, operated in delayed extraction and reflector modes, and α-hydroxycinnamic acid saturated in 0.1% trifluoroacetic acid in 70% acetonitrile as the matrix. A tryptic peptide mixture of *Cratylia floribunda* seed lectin (SwissProt accession code P81517) prepared and previously characterized in our laboratory was used as a mass calibration standard (mass range 450–3300 Da). For determination of isotope-averaged molecular masses, the instrument was operated in the linear mode using 3,5-dimethoxy-4-hydroxycinnamic acid (sinapinic acid) saturated in 70% acetonitrile and 0.1% trifluoroacetic acid as the matrix. The mass calibration standard consisted of a mixture of the following proteins, whose isotope-averaged molecular mass in daltons are given in parentheses: bovine insulin (5,734.6), *E. coli* thioredoxin (11,674.5), and horse apomyoglobin (16,952.6).

Protein concentration was determined with the bicinchoninic acid (BCA) protein quantification kit (Pierce) with bovine serum albumin as a standard, or by amino acid analysis (after hydrolysis in 6 N HCl for 24 h at 110 °C in air-evacuated and sealed ampoules) using a Biochrom (Amersham Biosciences) amino acid analyzer.

Cloning and Production of Recombinant r-Jerdostatin-Thioredoxin-His_6 Fusion Proteins—The jerdostatin cDNA coding for wild-type and R21K fragments were amplified by PCR using primers synthesized by Sigma-Genosys (Haverhill, UK). The forward primer was 5′-CGTG**C**-

[4] The abbreviations used are: r-jerdostatin; recombinant jerdostatin; HPLC, high performance liquid chromatography; MALDI-TOF, matrix-assisted laser-desorption ionization time-of-flight; MS, mass spectrometry.

CATGGATTGTACAACTGGACCATG-3′, which contained a NcoI restriction site (underlined) and the sequence coding for the first six residues of the protein. The reverse primer was 5′-GC***CTCGAG***TA***TTA***GCCATTCCCGGGATAAC-3′, which includes a restriction site for XhoI (underlined), a stop codon (in italics and bold), and the last six C-terminal residues of jerdostatin. The PCR protocol included denaturation at 94 °C for 2 min, followed by 40 cycles of denaturation (10 s at 94 °C), annealing (15 s at 55 °C), and extension (20 s at 72 °C), and a final extension for 7 min at 72 °C. The amplified fragments were purified using the Perfect Pre Gel Clean Up kit (Eppendorf, Hamburg, Germany) and cloned in a pGEM-T vector (Promega, Madison, WI). *E. coli* DH5α cells (Novagen, Madison, WI) were transformed by electroporation using an Eppendorf 2510 electroporator following the manufacturer's instructions. Positive clones, selected by growing the transformed cells in Luria broth (LB) medium containing 100 μg/ml ampicilin, were confirmed by PCR amplification using the above primers, and the PCR-amplified fragments were sequenced (using an Applied Biosystems model 377 DNA sequencer) to check the correctness of the sequences of the wild-type and the R21K jerdostatins open reading frame.

To construct an expression vector of jerdostatin-thioredoxin-His_6 wild-type and mutated fusion proteins the pGEM-T-jerdostatin plasmid and a pET32a vector (Novagen) were digested with NcoI and XhoI for 12 h at 37 °C and the 132-bp jerdostatin fragments and the pET32a vector were purified after agarose gel electrophoresis with the Eppendorf Perfect Pre Gel Clean Up kit. The jerdostatin fragments and the open pET32a vector were ligated with T4 DNA ligase (Invitrogen) overnight at 13 °C. These constructs were used to transform electrocompetent *E. coli* DH5α cells. The plasmidic DNAs from positive clones were used to transform (by electroporation) *E. coli* Origami® B cells (Novagen). Another pool of cells were transformed with mock pET32a plasmid and used as negative control for the recombinant expression of jerdostatin-thioredoxin fusion protein.

Recombinant Expression of Jerdostatin-Thioredoxin-His_6 Fusion Proteins—Positive *E. coli* Origami B clones, shown by PCR to contain the jerdostatin-thioredoxin fusion constructs, wild-type or R21K mutant, were grown overnight at 37 °C in LB medium containing 100 μg/ml of ampicillin, 33 μg/ml of kanamycin, and 12 μg/ml of tetracyclin, followed by a 1:10 (v/v) dilution in the same medium. For the induction of the expression of the recombinant fusion proteins, isopropyl β-D-thiogalactosidase was added to a final concentration of 1 mM, and the cell suspensions were incubation for another 7 h at 37 °C. Thereafter, the cells were pelleted by centrifugation, resuspended in the same volume of 20 mM sodium phosphate, 150 mM NaCl, pH 7.4, washed three times with the same buffer, and resuspended in 100 ml/liter of initial cell culture of 20 mM sodium phosphate, 250 mM NaCl, 10 mM imidazole, pH 7.4. The cells were lyzed by sonication (15 cycles of 15 s sonication followed by 1 min resting) in an ice bath. The lysates were centrifuged at 10,000 × *g* for 30 min at 4 °C, and the soluble and the insoluble fractions were analyzed by SDS-15% polyacrylamide gel electrophoresis.

Purification of Recombinant Jerdostatin Molecules—The jerdostatin-thioredoxin-His_6 fusion proteins, wild-type and R21K mutant, were purified from the soluble fraction of positive *E. coli* Origami clone, the lysate was purified by affinity chromatography using an ÄKTA Basic chromatograph equipped with a 5-ml HisTrap HP column (Amersham Biosciences) equilibrated in 20 mM sodium phosphate, 250 mM NaCl, 10 mM imidazole, pH 7.4, buffer. After absorbance at 280 nm of the flow-through fraction reached baseline, the bound material was eluted at a flow rate of 1.5 ml/min with a linear gradient of 10–500 mM imidazole for 60 min. The purified protein fractions (checked by SDS-PAGE) were pooled, dialyzed against 50 mM Tris/HCl, pH 7.4, and digested with 0.25 units of enterokinase (Invitrogen) per mg of recombinant protein. The reaction mixture was freed from enterokinase by chromatography on a 0.5-ml column of agarose-trypsin inhibitor (Sigma) equilibrated and eluted with 50 mM Tris-HCl, pH 7.4. Jerdostatin was separated from thioredoxin-His_6 by chromatography of the agarose-trypsin-inhibitor non-bound fraction on a HisTrap column (as above), and the non-bound and retarded fractions, both containing jerdostatin, were further purified by reverse-phase HPLC followed by size-exclusion chromatography using an ÄKTA Basic chromatograph equipped with a Superdex Peptide column (Amersham Biosciences) eluted with phosphate-buffered saline buffer at a flow rate of 0.3 ml/min. The purity of the isolated proteins was assessed by SDS-PAGE, reverse-phase HPLC, N-terminal sequence analysis (using an Applied Biosystems Procise instrument), and MALDI-TOF mass spectrometry as described above for the KTS-disintegrins, and nanoelectrospray ionization mass spectrometry using a QTrap instrument (Applied Biosystems) equipped with a nanoelectrospray source (Proxeon, Denmark). Protein concentration of purified recombinant jerdostatin was determined spectrophotometrically using an ϵ at 280 nm of 10,677 M^{-1} cm^{-1} calculated by amino acid analysis as above.

In-gel Tryptic Digestion and Mass Fingerprinting—The recombinant expression of the jerdostatin-thioredoxin-His_6 fusion proteins and the purification of r-jerdostatin molecules were monitored by SDS-PAGE and mass fingerprinting. To this end, SDS-PAGE separated polypeptides were subjected to automated digestion with sequencing grade bovine pancreatic trypsin (Roche) at a final concentration of 20 ng/μl of 50 mM ammonium bicarbonate, pH 8.3, using a ProGest digestor (Genomic Solutions) following the manufacturer's instructions. Digestions were done with prior reduction with dithiothreitol (10 mM for 15 min at 65 °C) and carbamidomethylation with iodoacetamide (50 mM for 60 min at room temperature). The tryptic peptide mixtures were freed from reagents using a C18 Zip-Tip pipette tip (Millipore) activated with 70% acetonitrile and equilibrated in 0.1% trifluoroacetic acid. Following protein adsorption and washing with 0.1% trifluoroacetic acid, the proteins were eluted with 3 μl of 70% acetonitrile and 0.1% trifluoroacetic acid. For mass fingerprinting analysis, 0.85 μl of the digests were spotted onto a MALDI-TOF sample holder, mixed with an equal volume of a saturated solution of α-cyano-4-hydroxycinnamic acid (Sigma) in 50% acetonitrile containing 0.1% trifluoroacetic acid, dried, and analyzed with an Applied Biosystems Voyager-DE Pro MALDI-TOF mass spectrometer, operated in delayed extraction and reflector modes, as above. The peptide mass fingerprint obtained from each electrophoretic band was compared with the expected proteolytic digest of the fusion protein using the program PAWS.[5]

Collision-induced Dissociation by Tandem Mass Spectrometry—For peptide sequencing, the protein digest mixture was subjected to electrospray ionization tandem mass spectrometric (MS/MS) analysis using a QTrap mass spectrometer (Applied Biosystems) equipped with a nanoelectrospray source (Protana, Denmark). Doubly charged ions selected after Enhanced Resolution MS analysis were fragmented using the Enhanced Product Ion with the Q_0 trapping option at 250 atomic mass units/s across the entire mass range. For MS/MS experiments, Q_1 was operated at unit resolution, the Q_1 to Q_2 collision energy was set to 35 eV, the Q_3 entry barrier was 8 V, the linear ion trap Q_3 fill time was 250 ms, and the scan rate in Q_3 was 1000 atomic mass units/s. Collision-induced dissociation spectra were interpreted manually or using the on-line form of the MASCOT program.[6]

[5] Proteometrics, available at prowl.rockefeller.edu.
[6] www.matrixscience.com.

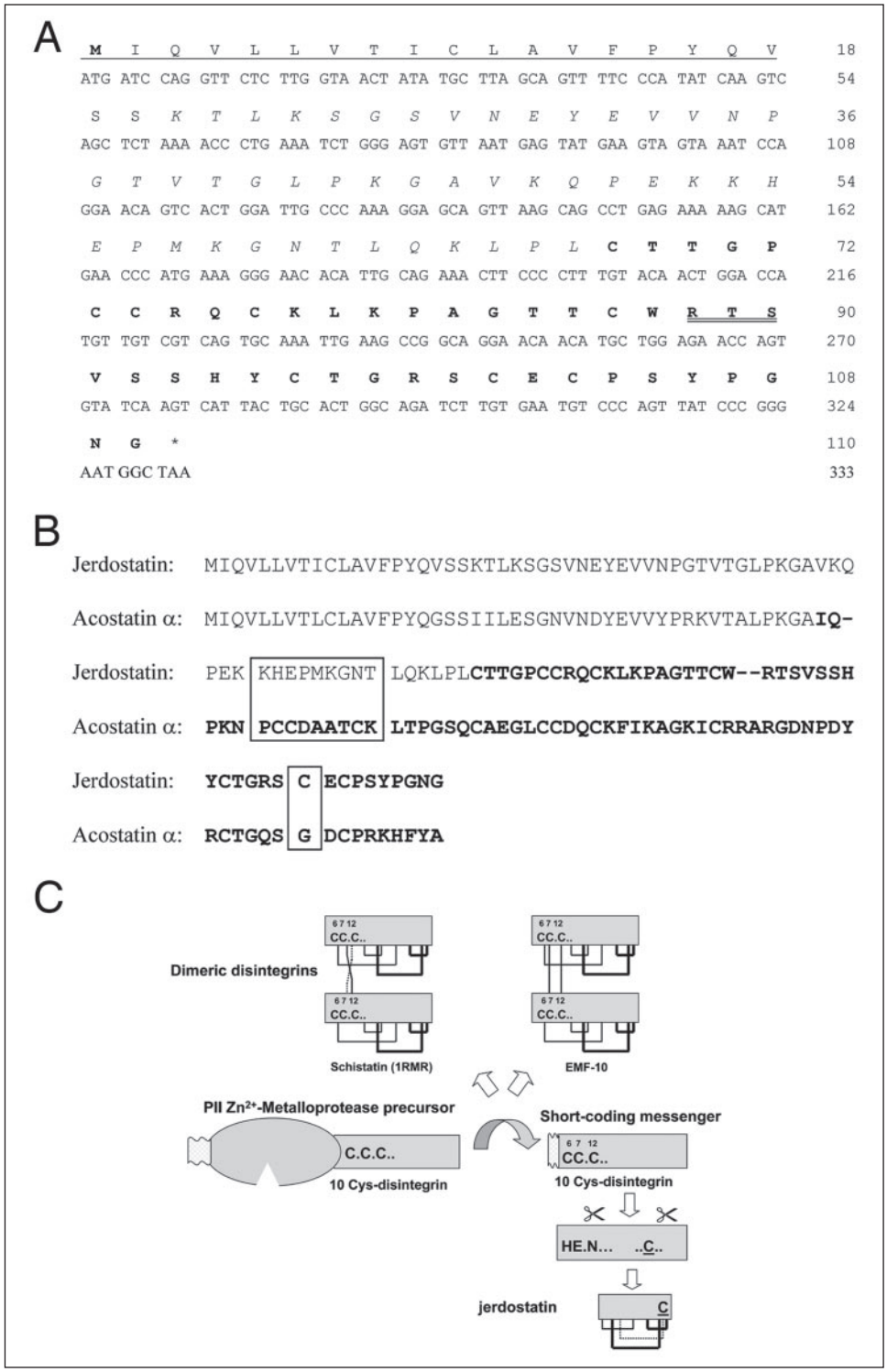


FIGURE 1. *A*, the complete nucleotide and deduced amino acid sequences of jerdostatin. The nucleotide sequence is numbered in the 5′to 3′ direction, from the initial codon ATG to the stop codon TAA. The signal sequence is *underlined*, and the pro-peptide and the short disintegrin domain are indicated in *italics* and *boldface*, respectively. The RTS integrin binding motif is *double underlined*. *B*, comparison of the cDNA-deduced amino acid sequences of jerdostatin and acostatin-α precursors. The disintegrin domains are in *boldface*. The regions of acostatin-α containing the three cysteine residues absent in jerdostatin, and the jerdostatin containing the short disintegrin-specific cysteine residue are *boxed*. *C*, cartoon depicting the proposed common ancestry of the messenger precursors coding for the short disintegrin jerdostatin and dimeric disintegrins. The proposed evolutionary pathway includes the removal of the metalloproteinase domain from a PII-metalloproteinase precursor gene. Key events in the emergence of jerdostatin appear to be the substitutions of the first three cysteine residues (Cys^6, Cys^7, and Cys^{12} in the dimeric disintegrin subunit precursor) by His, Glu, and Asn, respectively, impairing thereby dimerization through either homologous Cys^{A7}–Cys^{B12} and Cys^{A12}–Cys^{B7} linkages, as reported for Schistatin (44), or Cys^7–Cys^7 and Cys^{12}–Cys^{12}, as determined for EMF-10 (40); the appearance of a novel cysteine residue at position 101 (short-coding precursor numbering) between the 9th and 10th cysteine of the dimeric disintegrin subunit precursor (*C*) enabling the short disintegrin-specific disulfide bond is depicted by a *broken line*, and the proteolytic processing of the N- and C-terminal regions (*scissors*). The proposed disulfide bond pattern for jerdostatin is as determined for obtustatin (24). The two conserved disulfide bonds in the structures of dimeric disintegrin subunits and the short disintegrins are represented by *thick lines*.

Quantitation of Free Cysteine Residues and Disulfide Bonds—For quantitation of free cysteine residues and disulfide bonds (30), the purified proteins dissolved in 10 μl of 50 mM HEPES, pH 9.0, 5 M guanidine hydrochloride containing 1 mM EDTA) were heat denatured at 85 °C for 15 min, allowed to cool at room temperature, and incubated with either 10 mM iodoacetamide for 1 h at room temperature, or with 10 mM 1,4-dithioerythritol (Sigma) for 15 min at 80 °C, followed by addition of iodoacetamide at 25 mM final concentration and incubation for 1 h at room temperature. Carbamidomethylated proteins were freed from reagents using a C18 ZipTip pipette tip (Millipore) after activation with 70% acetonitrile and equilibration in 0.1% trifluoroacetic acid. Following protein adsorption and washing with 0.1% trifluoroacetic acid, the PE-proteins were eluted onto the MALDI-TOF plate with 1 μl of 70% acetonitrile and 0.1% trifluoroacetic acid and subjected to mass spectrometric analysis as above.

The number of free cysteine residues (N_{SH}) was determined using Equation 1,

$$N_{SH} = (M_{1A} - M_{NAT})/57.05 \quad \text{(Eq. 1)}$$

where M_{IA} is the mass of the denatured but nonreduced protein incubated in the presence of iodoacetamide; M_{NAT} is the mass of the native, HPLC-isolated protein; and 57.05 is the mass increment because of the carbamidomethylation of one thiol group.

The number of total cysteine residues (N_{Cys}) can be calculated from Equation 2,

$$N_{Cys} = (M_{CM} - M_{NAT})/57.05 \quad \text{(Eq. 2)}$$

where M_{CM} is the mass (in Da) of the reduced and carbamidomethylated protein.

Finally, the number of disulfide bonds $N_{S\text{-}S}$ can be calculated from Equation 3.

$$N_{S\text{-}S} = (N_{cys} - N_{SH})/2 \quad \text{(Eq. 3)}$$

Cell Adhesion Studies—Adhesion studies of cultured cells labeled with 5-chloromethyl fluorescein diacetate were performed essentially as described (31, 32). For inhibition studies, increasing concentrations of disintegrins were incubated for 30 min at 37 °C with 1×10^5 5-chloromethyl fluorescein diacetate-labeled cells in the wells of a 96-well enzyme-linked immunosorbent assay plate, previously covered with collagen IV (2 μg/ml in 100 μl of Hank's balanced salt solution containing 3 mM Mg^{2+}). After washing with the same buffer, the adhered cells were lysed with Triton X-100 and the plate was read using a FLx800 fluorescence plate reader. The percentage inhibition was calculated by comparison with the fluorescence values obtained from control samples without integrin inhibitors.

NMR Spectroscopy—For one-dimensional ^{1}H NMR analyses, each of the two HPLC fractions of wild-type r-jerdostatin was dissolved in 5% D_2O, 95% H_2O and placed in 5-mm Shigemi H_2O/D_2O susceptibility matched NMR tubes. Final concentrations of the samples were determined by UV-visible spectroscopy and were 2.4 and 1.4 mM for conformers HPLC-1 and HPLC-2, respectively. NMR spectra were recorded on a Bruker Avance NMR spectrometer operating at a ^{1}H frequency of 500.13 MHz and equipped with conventional BBI dual ^{1}H-Broadband probe. The spectra were processed on a SGI O_2 work station running the XWIN-NMR version 3.1 software. Processing included stages of apodization with a Gaussian function and zero filling to double number of points. Solvent signal suppression was achieved either by using the double WATERGATE pulse sequence (33, 34) or by low-power irradiation at the water resonance frequency. A recycling relaxation delay of 1.5 segments between transients was employed in all experiments. Spectra recorded for both samples included one-dimensional double WATERGATE ^{1}H and one-dimensional ^{1}H NOE.

RESULTS AND DISCUSSION

The Structure of the Jerdostatin Open Reading Frame Provides Clues for Its Evolutionary Emergence—Jerdostatin represents a novel non-RGD short disintegrin encoded by a cDNA amplified from the venom gland mRNA of *T. jerdonii* by reverse transcriptase-PCR using primers complementary of the highly conserved 5′-and 3′-noncoding regions of other *Trimeresurus* disintegrins genes. The cDNA of jerdostatin comprised 369 bp (GenBank accession code AY262730) coding for an open reading frame of 333 bp including a signal sequence (1–20), a pre-peptide (21–68), and an obtustatin-like short disintegrin domain (residues 69–110) (Fig. 1*A*).

Although the vast majority of disintegrins, including all known monomeric PIII, long, and medium-sized disintegrins, are derived by proteolysis of a large mosaic metalloproteinase precursor (35), the α-subunit of the dimeric disintegrin acostatin from *Agkistrodon contortrix contortrix* venom has been reported to be coded for by a short-coding mRNA (36) similar to the jerdostatins messenger (Fig. 1*B*). Noteworthy, the jerdostatin pre-peptide sequence encompasses a region that is a homolog of the N-terminal sequence of acostatin-α harboring the first 3 cysteine residues of the mature molecule. Current biochemical and genetic data support the view that the different groups of the disintegrin family evolved from a common ancestor and that structural diversification occurred through disulfide bond engineering (35). In line with this view, the jerdostatin cDNA structure reported here, which represents the first complete open reading frame of a short disintegrin, points to a mechanism for the emergence of jerdostatin from a short-coding gene similar to that of acostatin-α. Fig. 1*C* depicts a scheme of the proposed evolutionary pathway, which involves substitutions of the first three N-terminal cysteine residues, the appearance of the short disintegrin-specific cysteine at the C-terminal region (*underlined*), and proteolytic processing of the precursor molecule at the N- and C-terminal regions. It is worth noting that the known native -fold of short disintegrins adopt a slightly different disulfide bond pattern than that of the dimeric disintegrin chains (Fig. 1*C*), providing further possibilities for the evolution of the structure and function of this family of integrin antagonists.

Recombinant Expression of Two HPLC Conformers of Jerdostatin—The deduced primary structure of jerdostatin exhibits 80–85% amino acid sequence identity with the KTS-disintegrins lebestatin from *M. lebetina*, obtustatin from *V. lebetina obtusa* venom (11, 21), and viperistatin isolated from the venom of *V. palestinae* (22) (Fig. 2). Noteworthy, the 7–9 different residues between jerdostatin and the KTS-disintegrins are segregated within the C-terminal half of the molecule, including the

		$IC_{50\%}$ (nM)
	-3 -1 1 5 10 15 20 25 30 35 40	
r-jerdostatin-R21	amdCTTGPCCRQCKLKPAGTTCW*R*T*S*VSSHYCTGRSCECPSYPGNG	180
r-jerdostatin-K21	amdCTTGPCCRQCKLKPAGTTCW*K*T*S*VSSHYCTGRSCECPSYPGNG	703
Obtustatin	CTTGPCCRQCKLKPAGTTCW*K*T*S*LTSHYCTGKSCDCPLYPG	2
Lebestatin	CTTGPCCRQCKLKPAGTTCW*K*T*S*RTSHYCTGKSCDCPSYPG	0.4
Viperistatin	CTTGPCCRQCKLKPAGTTCW*K*T*S*RTSHYCTGKSCDCPVYQG	0.08

FIGURE 2. **Comparison of the amino acid sequences and $\alpha_1\beta_1$-collagen IV inhibitory activities of r-jerdostatin-R21, r-jerdostatin-K21, obtustatin (*V. lebetina obtusa* (21)), lebestatin (*M. lebetina*), and viperistatin (*V. palestinae* (22)).** Cysteine residues are *underlined* and the active motifs are in *italics*. Amino acid residues that are different from viperistatin are *double underlined*. Residues −3 and −1 correspond to the expression vector pET32a. Data not shown for lebstatin (L. Sanz *et al.*, manuscript in preparation).

integrin-recognition loop and the C-terminal tail, two structural elements that form a continuous functional epitope in the three-dimensional structure of obtustatin (24, 25). In particular, jerdostatin contains a novel RTS motif instead of the KTS tripeptide found in lebestatin, obtustatin, and viperistatin. The KTS motif has been shown to endow disintegrins with selective inhibitory activity of the *in vitro* adhesion of integrin $\alpha_1\beta_1$ to immobilized collagen IV (21) and of angiogenesis *in vivo* (11). To investigate the biological activity of jerdostatin, the wild-type disintegrin was expressed in *E. coli* Origami B cells as a jerdostatin-thioredoxin-His_6 fusion protein. The Origami B cells are derived from a *lacZY* mutant of *E. coli* BL21 and provide mutations in both the thioredoxin reductase (*trxB*) and glutathione reductase (*gor*) genes, greatly enhancing disulfide bond formation in the bacterial cytoplasm (37). Induction of the expression of the recombinant fusion protein construct was independent of the addition of the (0.1 mM) Lac operon inducer isopropyl β-D-thiogalactosidase, and the recombinant fusion protein was produced in approximately equal amounts in the soluble and insoluble cell lysate fractions (Fig. 3, *lanes c* and *d*). The expression of the fusion protein in these two subcellular fractions was assessed by MALDI-TOF mass fingerprinting of in-gel tryptic digests followed by collision-induced dissociation of selected monoisotopic ions. In particular, the simultaneous presence of ions from thioredoxin (903.3 (2^+), MIAPILDEIADEYQGK, and 634.3 (2^+), LNIDQNPGTAPK) and jerdostatin (595.4 (2^+) LKPAGTTCWR, and 627.9 (2^+), TSVSSHYCTGR) demonstrated that the protein band corresponded to the expected recombinant fusion protein product.

Affinity chromatography on a HisTrap column of the enterokinase degradation mixture of the jerdostatin-thioredoxin-His_6 fusion protein yielded major (80%) non-bound and minor (20%) retarded fractions. Both protein fractions eluted at the same position from the Superdex Peptide size-exclusion column used to complete the purification protocol, exhibited distinct reverse-phase HPLC elution profile, and had the same amino acid sequence (AMDCTTGPCCRQCKLKP . . .) (Fig. 2), MALDI-TOF native isotope-averaged molecular mass (4898.6 Da) (Fig. 3*B*), and tryptic peptide mass fingerprinting expected for reduced and carbamidomethylated r-jerdostatin. Sequence analysis of the tryptic peptides was done by collision-induced fragmentation tandem mass spectrometry and confirmed the MALDI-TOF mass fingerprint assignments. The final purification yields of the two jerdostatin isoforms, designated according to their elution order from the reverse-phase HPLC column as conformers-1 and -2 of wild-type r-jerdostatin (r-jerdostatin-R21), were about 2 and 0.5 mg, respectively, per liter of Origami B cell culture. The monoisotopic molecular masses of r-jerdostatin-R21 conformers 1 and 2, measured by nanoelectrospray ionization mass spectrometry, were both 4894.6 ± 0.3 Da (Fig. 3*B*, *inset*), which matched accurately the calculated value for the cDNA-derived amino acid sequence of the disintegrin (Fig. 2) with fully oxidized cysteine residues (calculated mass 4894.8 Da). Furthermore, mass spectrometric analysis of the reduced and carbamidomethylated conformers 1 and 2 yielded the same isotope-averaged molecular mass of 5363.2 Da. Incubation of r-jerdostatin-R21 isoforms 1 and 2 with iodoacetamide under denaturing but nonreducing conditions did not change their molecular masses. Hence, the mass difference of 464.6 Da between the native and reduced and carbamidomethylated proteins clearly indicated that each r-jerdostatin isoform contained eight cysteine residues engaged in the formation of 4 disulfide bonds. Taken together, these data, along with the different behaviors of r-jerdostatin-R21 conformers 1 and 2 on reverse-phase HPLC suggested that the two isoforms may represent structural conformers of r-jerdostatin. The stronger binding of conformer 2 to the C18 matrix indicates that r-jerdostatin conformer 2 exposes more

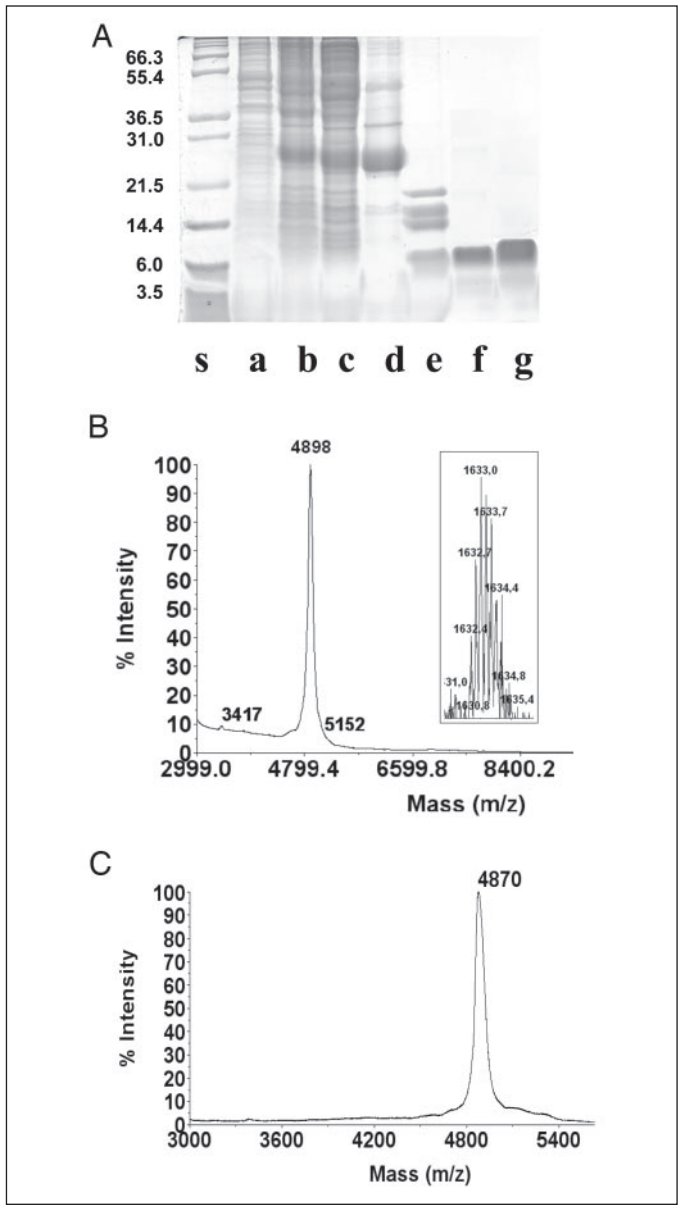


FIGURE 3. *A*, SDS-15% polyacrylamide gel electrophoretic analysis of the expression and purification of wild-type (R21) r-jerdostatin. *Lane a*, cell lysates of *E. coli* Origami B cells transformed with the mock pET32a plasmid. *Lanes b* and *c*, 100 μg of total proteins of the insoluble and soluble fractions, respectively, of cell lysates of Origami B cells expressing the jerdostatin-thioredoxin fusion protein construct. *Lane d*, HisTrap affinity purified jerdostatin-thioredoxin fusion protein. *Lane e*, protein mixture generated by digestion with enterokinase of the jerdostatin-thioredoxin fusion protein. *Lanes f* and *g*, r-jerdostatin-1 and r-jerdostatin-2 proteins purified by reverse-phase chromatography of the flow-through and the retarded fractions, respectively, of HisTrap affinity chromatography of the protein mixture of *lane e*. *Lane S*, molecular mass markers (Mark12®, Invitrogen), from top to bottom: bovine serum albumin (66.3 kDa), glutamic dehydrogenase (55.4 kDa), lactate dehydrogenase (36.5 kDa), carbonic anhydrase (31.0 kDa), soybean trypsin inhibitor (21.5 kDa), lysozyme (14.4 kDa), aprotinin (6 kDa), and insulin B-chain (3.5 kDa). *B*, MALDI-TOF (linear mode) mass spectrum of purified HPLC conformer 1 of r-jerdostatin-R21. The *inset* shows the triply charged isotope cluster of r-jerdostatin-R21 determined by nanoelectrospray mass spectrometry. The same results were obtained with the HPLC conformer 2 of r-jerdostatin-R21. *C*, isotope-averaged molecular mass of purified HPLC conformer 1 of r-jerdostatin-K21 determined by MALDI-TOF mass spectrometry. The same result was obtained with the HPLC conformer 2 of r-jerdostatin-K21.

hydrophobic surface than conformer 1. We sought to investigate the possibility that alternative disulfide bond connectivities could account for the different chromatographic behaviors of the two r-jerdostatin-R21 conformers. Determination of their disulfide bond pattern(s) was not possible, however, because both proteins were resistant to enzy-

matic proteolysis and chemical degradation with oxalic acid. Degradation with oxalic acid has been previously used to assign disulfide bonds in echistatin (38) (short, RGD-containing disintegrin), bitistatin (39) (long RGD-disintegrin), and EMF10 (40) (dimeric disintegrin), but did not work with obtustatin, indicating that the KTS/RTS-disintegrins are unique among disintegrins regarding their unusual stable conformation.

Conformers 1 and 2 Represent, Respectively, Native and Unfolded r-Jerdostatin-R21 Molecular Species—The folding status of conformers 1 and 2 of r-jerdostatin-R21 was assessed by one-dimensional ^{1}H NMR (41, 42). NMR spectra of the two samples showed a good signal-to-noise ratio, and water signal line width in the spectra of both samples, even without water suppression, was very narrow, indicating a good field homogeneity. Spectra recorded using the double WATERGATE solvent signal suppression method are shown in Fig. 4. The narrow proton peaks along with good resonance dispersion of at least 50 different peaks in the amide and aromatic proton regions (Fig. 4*A*) clearly indicate that conformer 1 possesses a well folded structure. On the other hand, the NMR spectrum of conformer 2 shown in Fig. 4*B* displays low resonance dispersion and broad peaks, which are indications of an unfolded flexible protein. Spectral differences between both r-jerdostatin-R21 conformers are particularly dramatic in the methyl proton region, from 0.5 to 1.5 ppm, where the dispersed and narrow set of peaks of conformer 1 contrasts with the presence of a wide band at the 1 ppm position in conformer 2. It is also worth noting that the peak at 10.02 ppm, which belongs to the HN side chain of the single r-jerdostatin tryptophan residue at position 20 (Fig. 2), remains at the same position in the spectra of both r-jerdostatin conformers, suggesting that this proton is exposed to the solvent and does not participate in intra-protein interactions. However, the significant broadening experimented by this proton peak in the spectrum of conformer 2 (Fig. 4*B*) indicates a larger degree of flexibility of Trp^{20} in the unfolded *versus* the folded (conformer 1) r-jerdostatin-R21 species.

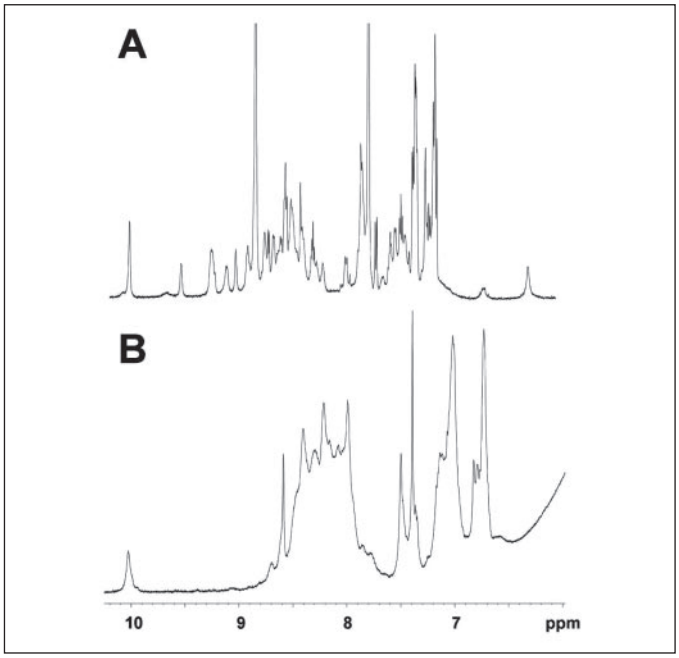


FIGURE 4. **One-dimensional ^{1}H double WATERGATE NMR spectra recorded in 95% H_2O, 5% D_2O at 500 MHz and 25°C showing the amide and aromatic proton region narrow proton peaks of HPLC conformers 1 (*A*) and 2 (*B*) of r-jerdostatin-R21.** The good resonance dispersion in *panel A*, characteristic of a well folded protein, contrasts with the low resonance dispersion and broad peaks in *panel B*, indicative of an unfolded flexible protein.

Conformers 1 and 2 of r-Jerdostatin-R21 Exhibit the Same Integrin Inhibitory Specificity Although Distinct Potency—Conformers 1 and 2 of r-jerdostatin-R21 were screened against a panel of integrins using the same cell adhesion inhibition assays described for the KTS-disintegrins obtustatin (11, 21) and viperistatin (22). Both r-jerdostatin conformers proved to be selective inhibitors of the binding of the $\alpha_1\beta_1$ integrin to collagen IV (TABLE ONE, Fig. 5), and none of them showed inhibitory activity toward other integrins such as $\alpha_{IIb}\beta_3$, $\alpha_v\beta_3$, $\alpha_2\beta_1$, $\alpha_5\beta_1$ $\alpha_4\beta_1$, $\alpha_6\beta_1$, and $\alpha_9\beta_1$ (TABLE ONE). The restricted integrin specificity of conformers 1 and 2 was highlighted by the fact that neither r-jerdostatin-R21 species blocked the adhesion of α_2-K562 cells to collagen ligands (TABLE ONE), a functional featured shared by the KTS-disintegrins. However, the IC_{50} of conformer 1 (180 nM) was, respectively, about 90, 900, and 2250 times less potent than obtustatin (IC_{50} 2 nM), lebestatin (IC_{50} 0.2 nM), and viperistatin (IC_{50} 0.08 nM), inhibiting the binding of cells expressing integrin $\alpha_1\beta_1$ to immobilized collagen IV. On the other hand, r-jerdostatin-R21 conformer 2 (IC_{50} 5950 nM) was 33 times less active than conformer 1 (Fig. 5, TABLE ONE). These functional data in conjunction with the one-dimensional ^{1}H NMR analysis of the two r-jerdostatin-R21 conformers discussed above and shown in Fig. 4 supports the view that conformer 1 has a native, fully active dis-

TABLE ONE

Inhibitory effects of the HPLC-1 and HPLC-2 conformers of wild-type r-jerdostatin-R21 and the mutant r-jerdostatin-K21 on various integrins in cell adhesion assays

The data represent the mean of three experiments. Coll, collagen; LM, laminin; FN, fibronectin; VCAM-1, vascular cell adhesion molecule-1; FG, fibrinogen; VN, vitronectin. α_1-, α_2-, and α_6-K562, K562 cells transfected with α_1, α_2, or α_6 integrins; SW480α 9, SW480 cells transfected with the α_9 integrin. HPLC-1 and HPLC-2, conformers 1 and 2, respectively.

Cell	Integrin	Ligand	IC_{50}			
			r-Jerdostatin-R21		r-Jerdostatin-K21	
			HPLC-1	HPLC-2	HPLC-1	HPLC-2
			nM		*nM*	
α_1-K562	$\alpha_{1\beta1}$	CollIV	180	5,950	703	12,500
α_2-K562	$\alpha_{2\beta1}$	Coll I	>24,000	>24,000	>24,000	>24,000
α_2-K562	$\alpha_{2\beta1}$	CollIV	>24,000	>24,000	>24,000	>24,000
α_6-K562	$\alpha_{6\beta1}$	LM	>24,000	>24,000	>24,000	>24,000
K562	$\alpha_{5\beta1}$	FN	>24,000	>24,000	>24,000	>24,000
Jurkat	$\alpha_{4\beta1}$	VCAM-1	>24,000	>24,000	>24,000	>24,000
SW480α9	$\alpha_{9\beta1}$	VCAM-1	>24,000	>24,000	>24,000	>24,000
Platelets	$\alpha_{IIb\beta3}$	FG	>24,000	>24,000	>24,000	>24,000
JY	$\alpha_{v\beta3}$	VN	>24,000	>24,000	>24,000	>24,000

integrin-fold, whereas conformer 2 may represent a non-native, activity compromised disintegrin molecule.

The lower inhibitory activity of the RTS *versus* KTS-disintegrins suggests that the amino acid residues that differentiate jerdostatin from obtustatin/viperistatin may create a distinct chemical environment responsible for its decreased potency, although these differences do not affect the restricted inhibitory selectivity of jerdostatin toward integrin $\alpha_1\beta_1$. Among them, Arg[21], Val[24], and Ser[25] belong to the integrin binding loop, Arg[32] and Glu[35] lay at the face opposite to the integrin binding loop, Ser[38] forms part of the hydrophobic core of the protein, and Pro[40], Asn[42], and Gly[43] reside in the C-terminal region of the molecule (Fig. 6). Mutations at positions 32, 35, and 38 may not significantly alter the conformation of the disintegrin, and may therefore represent neutral mutations. In line with this assumption, pairwise comparison of the amino acid sequences and inhibitory potency of the KTS-disintegrins shown in Fig. 2 indicate that substitution at positions 38 and 40 impair the potency of lebestatin *versus* viperistatin. An extra mutation R24L in obtustatin further decreases 1 order of magnitude the $\alpha_1\beta_1$ blocking activity of this disintegrin when compared with lebestatin. In agreement with this reasoning, comparison of the $\alpha_1\beta_1$ inhibitory activities of viperistatin and obtustatin, using synthetic peptides representing their integrin binding loops (viperistatin, ^{19}CWKTS**R**TSHYC^{29}; obtustatin, ^{19}CWKTS**L**TSHYC^{29}), showed that the 25-fold increased inhibitory activity of viperistatin over obtustatin was because of an Arg/Leu mutation at position 24 of the integrin binding loop and a Gln/Pro substitution at position 40 of the C-terminal region (22).

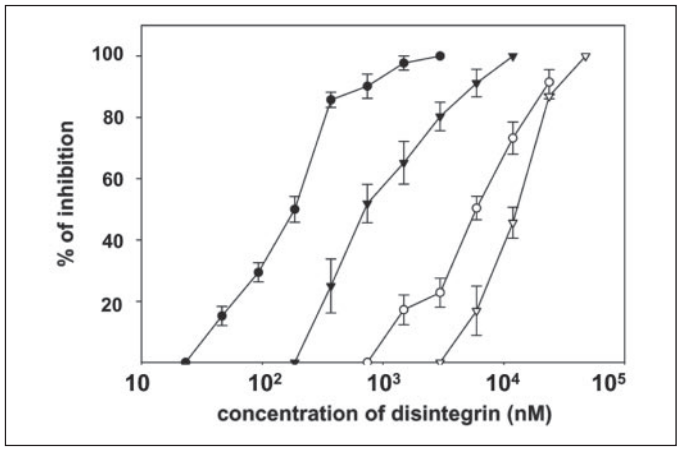


FIGURE 5. **Effect of the native folded conformer 1 (*filled circles*) and the unfolded conformer 2 (*open circles*) of r-jerdostatin-R21 and conformer 1 (*filled triangles*) and conformer 2 (open triangles) of r-jerdostatin-K21 on the adhesion of α_1-K562 cells to immobilized collagen IV.** *Error bars* represent S.D. from three duplicated experiments.

The possible contribution of integrin binding loop residues Arg[21], Val[24], and Ser[25] to the decreased $\alpha_1\beta_1$ inhibitory ability of r-jerdostatin-R21 *versus* its homologue KTS-disintegrins was assessed using 9 sets of positional-restricted combinatorial synthetic peptides. Each set contained 19 peptides representing the entire integrin-recognition loop of obtustatin but differing in the residue at a single position (^{19}CX_1KTSLTSHYC^{29}; ^{19}CWX_2TSLTSHYC^{29}; etc., where X_n is an equimolar mixture of all amino acids except cysteine). Compared with an obtustatin control loop peptide, sets X_2, X_3, X_4, and X_7, exhibited about 5-fold enhanced activity, whereas X_1, X_5, X_6, X_8, and X_9 showed 2–5-fold decreased activity. Although the neat differences in activity were modest, probably because of compensatory effects, the results were recurrent and converged to indicate that most of the integrin binding loop positions may play a functional role either through direct interactions with the receptor, or indirectly by maintaining the active conformation of the loop, as shown for Thr[22] of obtustatin (25). According to the NMR structure of obtustatin, residues at positions X_1, X_2, X_4, and X_5, are surface-exposed amino acids of the integrin binding loop. Noteworthy, X_2 corresponds to Lys[21], and thus the results indicating that substitutions at this position enhanced the $\alpha_1\beta_1$ inhibitory ability of the peptides suggests that KTS may be a suboptimal $\alpha_1\beta_1$ inhibitory motif. On the other hand, this result provided circumstantial evidence against a more potent inhibitory activity of KTS *versus* RTS. To further check this hypothesis, a single R21K mutant (r-jerdostatin-K21) was generated by site-directed mutagenesis, as described under "Experimental Procedures."

The r-Jerdostatin-K21 Mutant Is a Weaker Inhibitor of Integrin $\alpha_1\beta_1$ Than Wild-type r-Jerdostatin-R21—Recombinant jerdostatin exhibiting the integrin-binding motif ^{21}KTS^{23} instead of ^{21}RTS^{23}, r-jerdostatin-K21, was expressed and purified as described for wild-type r-jerdostatin-R21. Similar to wild-type disintegrin, two protein fractions differing in their HisTrap and reverse-phase HPLC elution times, but displaying the same amino acid sequence and the expected molecular mass of the fully disulfide-bonded mutant protein (4870 Da) (Fig. 3*C*), were purified. Furthermore, both conformers of the Lys[21] mutant selectively impaired the adhesion of α_1-K562 cells to collagen IV (TABLE ONE), although conformer 1 was significantly more potent than conformer 2. Nonetheless, the r-jerdostatin-K21 conformers are weaker inhibitors than the homologous wild-type proteins (TABLE ONE), indicating that the KTS motif is a less potent antagonist of the integrin $\alpha_1\beta_1$ than RTS.

Concluding Remarks—Based on its structural and functional characteristics, we propose that jerdostatin belongs, together with obtustatin, viperistatin, and lebestatin to the novel class of short-sized $\alpha_1\beta_1$-specific disintegrins. A distinct feature of jerdostatin is its novel ^{21}RTS^{23} motif, which appears to represent a more potent inhibitor of integrin $\alpha_1\beta_1$

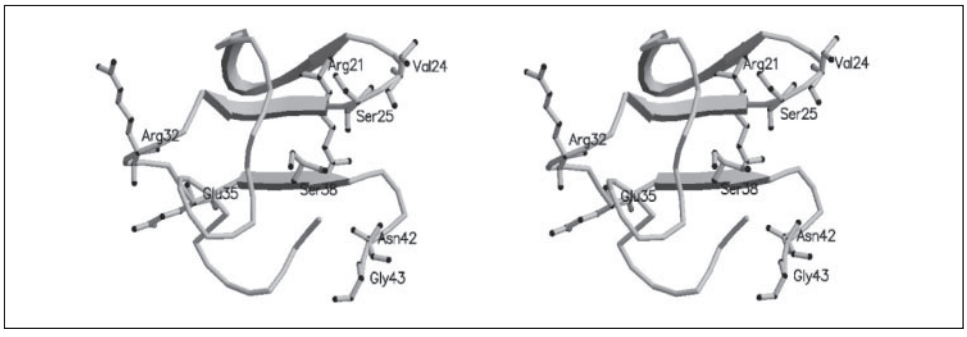


FIGURE 6. **Stereo drawing of a molecular model of wild-type r-jerdostatin-R21 based on the NMR solution structure of obtustatin (Protein Data Bank 1MPZ) (24) showing the location of residues (rendered in the ball-and-stick representation) that differentiate r-jerdostatin-R21 from obtustatin: Arg[21], Val[24], and Ser[25] within the integrin binding loop, Arg[32] and Glu[35] at the face opposing the integrin binding loop, Ser[38] in the hydrophobic core of the protein, and Pro[40], Asn[42], and Gly[43] in the C-terminal region of the molecule.** The conformation and relative orientation of C-terminal residues Asn[42] and Gly[43], which are absent in obtustatin, has been arbitrary modeled.

than KTS. The fact that recombinant wild-type jerdostatin is less active than KTS-disintegrins obtustatin, viperistatin, and lebestatin isolated from their natural sources, suggests that amino acid residues of jerdostatin outside of the integrin binding motif and departing from the primary structures of the KTS-disintegrins may create a distinct chemical environment responsible for the lower inhibitory activity of jerdostatin, although these substitutions do not affect the restricted inhibitory selectivity of jerdostatin toward integrin $\alpha_1\beta_1$. On the other hand, NMR studies have revealed that the integrin binding loop and the C-terminal tail of obtustatin (24, 25) and echistatin (43) are structurally linked and display concerted motions in the 100–300-ps time scale, strongly indicating that these two functional regions may form a conformational epitope engaged in extensive interactions with the target integrin receptor. Jerdostatin, like lebestatin, possesses serine and proline at positions 38 and 40, respectively, making it unlikely that these residues account for the decreased functional activity of the recombinant disintegrin. However, a distinct feature of r-jerdostatin is the presence of two C-terminal residues (^{42}NG^{43}), which in all venom-isolated members of the $\alpha_1\beta_1$-specific short disintegrins are post-translationally removed. The structural and functional consequences of the lack of C-terminal processing deserve further detailed investigation.

Acknowledgments—We thank the SCSIE of the University of Valencia for providing access to the NMR facility. We are grateful to Profs. Naziha Marrakchi and Mohamed El Ayeb (Institut Pasteur of Tunis) for sharing structural and functional data of lebestatin before publication.

REFERENCES

1. Ridley, A. J., Schwartz, M. A., Burridge, K., Firtel, R. A., Ginsberg, M. H., Borisy, G., Parsons, J. T., and Horwitz, A. R. (2003) *Science* **302,** 1704–1709
2. Springer, T. A., and Wang, J. (2004) *Adv. Protein Chem.* **68,** 29–63
3. Danen, E. H. (2005) *Curr. Pharm. Des.* **11,** 881–891
4. Rehn, M., Veikkola, T., Kukk-Valdre, E., Nakamura, H., Ilmonen, M., Lombardo, C. R., Pihlajaniemi, T., Alitalo, K., and Vouri, K. (2001) *Proc. Natl. Acad. Sci. U. S. A.* **98,** 1024–1029
5. Jin, H., and Varner, J. (2004) *Br. J. Cancer.* **90,** 561–565
6. Mistry, A., Harbottle, R., Hart, S., and Hodivala-Dilke, K. M. (2004) *Curr. Opin. Mol. Ther.* **5,** 603–610
7. Hynes, R. O. (2002) *Nat. Med.* **8,** 918–921
8. Hodivala-Dilke, K. M., Reynolds, A. R., and Reynolds, L. E. (2003) *Cell Tissue Res.* **314,** 131–144
9. Senger, D. R., Perruzzi, C. A., Streit, M., Koteliansky, V. E., De Fougerolles, A. R., and Detmar, M. (2002) *Am. J. Pathol.* **160,** 195–204
10. Senger, D. R., Claffey, K. P., Benes, J. E., Perruzzi, C. A., Sergiou, A. P., Detmar, M. (1997) *Proc. Natl. Acad. Sci. U. S. A.* **94,** 13612–13617
11. Marcinkiewicz, C., Wainreb, P. H., Calvete, J. J., Kisiel, D. G., Mousa, S. A., Tuszynski, G. P., and Lobb, R. R. (2003) *Cancer Res.* **63,** 2020–2023
12. Pozzi, A., Moberg, P. E., Miles, L. A., Wagner, S., Soloway, P., and Gardner, H. A. (2000) *Proc. Natl. Acad. Sci. U. S. A.* **97,** 2202–2207
13. Mercurio, A. M. (2002) *Am. J. Pathol.* **161,** 3–6
14. DeClerck, Y. A., Mercurio, A. M., Stack, M. S., Chapman, H. A., Zutter, M. M., Muschel, R. J., Raz, A., Matrisian, L. M., Sloane, B. F., Noel, A., Hendrix, M. J., Coussens, L., and Padarathsingh, M. (2004) *Am. J. Pathol.* **164,** 1131–1139
15. White, D. J., Puranen, S., Johnson, M. S., and Heino, J. (2004) *Int. J. Biochem. Cell Biol.* **36,** 1405–1410
16. Dickeson, S. K., Mathis, N. L., Rahman, M., Bergelson, J. M., and Santoro, S. A. (1999) *J. Biol. Chem.* **274,** 32182–32191
17. Tulla, M., Pentikäinen, O. T., Viitasalo, T., Käpylä, J., Impola, U., Nykvist, P., Nissinen, L., Johnson, M. S., and Heino, J. (2001) *J. Biol. Chem.* **276,** 48206–48212
18. Klekotka, P. A., Santoro, S. A., Ho, A., Dowdy, S. F., and Zutter, M. (2001) *Am. J. Pathol.* **159,** 983–992
19. Wang, R., Kini, R. M., and Chung, M. C. M. (1999) *Biochemistry* **38,** 7584–7595
20. Marcinkiewicz, C., Lobb, R. R., Marcinkiewicz, M. M., Daniel, J. L., Smith, J. B., Dangelmaier, C., Weinreb, P. H., Beacham, D. A., and Niewiarowski, S. (2000) *Biochemistry* **39,** 9859–9867
21. Moreno-Murciano, P., Daniel Monleón, D., Calvete, J. J., Celda, B., and Marcinkiewicz, C. (2003) *Protein Sci.* **12,** 366–371
22. Kisiel, D. G., Calvete, J. J., Katzhendel, J., Fertala, A., Lazarovici, P., and Marcinkiewicz, C. (2004) *FEBS Lett.* **577,** 478–482
23. Horii, K., Okuda, D., Morita, T., and Mizuno, H. (2004) *J. Mol. Biol.* **341,** 519–527
24. Moreno-Murciano, M. P., Monleón, D. Marcinkiewicz, C., Calvete, J. J., and Celda, B. (2003) *J. Mol. Biol.* **329,** 135–145
25. Monleón, D., Moreno-Murciano, M. P., Kovacs, H., Marcinkiewicz, C., Calvete, J. J., and Celda, B. (2003) *J. Biol. Chem.* **278,** 45570–45576
26. Yokosaki, Y., Palmer, E. L., Prieto, A. L., Crossin, K. L., Bourdon, M. A., Pytela, R., and Sheppard, D. (1994) *J. Biol. Chem.* **269,** 26691–26696
27. Yamada, D., Shin, Y., and Morita, T. (1999) *FEBS Lett.* **451,** 299–302
28. Pastor, M. T., De la Paz, M. L., Lacroix, E., Serrano, L., and Pérez-Payá, E. (2002) *Proc. Natl. Acad. Sci. U. S. A.* **99,** 614–619
29. Gill, S. C., and Hippel, P. H. (1989) *Anal. Biochem.* **182,** 319–326
30. Juárez, P., Sanz, L., and Calvete, J. J. (2004) *Proteomics* **4,** 327–338
31. Marcinkiewicz, C., Calvete, J. J., Marcinkiewicz, M. M., Raida, M., Lobb, R. R., Senadhi, V.-K., Huang, Z., and Niewiarowski, S. (1999) *J. Biol. Chem.* **274,** 12468–12473
32. Marcinkiewicz, C., Senadhi, V. K., McLane, M. A., and Niewiarowski, S. (1997) *Blood* **90,** 1565–1575
33. Piotto, M., Sandek, V., and Sklenar, V. (1992) *J. Biomol. NMR* **2,** 661–670
34. Liu, M., Mao, M., He, C., Huang, H., Nicholson, J. K., and Lindon, J. C. (1998) *J. Magn. Reson.* **132,** 125–129
35. Calvete, J. J., Moreno-Murciano, M. P., Theakston, R. D. G., Kisiel, D. G., and Marcinkiewicz, C. (2003) *Biochem. J.* **372,** 725–734
36. Okuda, D., Nozaki, C., Sekiya, F., and Morita, T. (2001) *J. Biochem.* (*Tokyo*) **129,** 615–620
37. Lobel, L., Pollak, S., Lutsbader, B., Klein, J., and Lutsbader, J. W. (2002) *Protein Expression Purif.* **25,** 124–133
38. Bauer, M., Sun, Y., Degenhardt, C., and Kozikowski, B. (1993) *J. Protein Chem.* **12,** 759–764
39. Calvete, J. J., Schrader, M., Raida, M., McLane, M. A., Romero, A., and Niewiarowski, S. (1997) *FEBS Lett.* **416,** 197–202
40. Calvete, J. J., Jürgens, M., Marcinkiewicz, C., Romero, A., Schrader, M., and Niewiarowski, S. (2000) *Biochem. J.* **345,** 573–581
41. Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley-Interscience, New York
42. Dyson, H. J., and Wright, P. E. (1996) *Annu. Rev. Phys. Chem.* **47,** 369–395
43. Monleón, D., Esteve, V., Kovacs, H., Calvete, J. J., and Celda, B. (2005) *Biochem. J.* **385,** 57–66
44. Bilgrami, S., Tomar, S., Yadav, S., Kaur, P., Kumar, J., Jabeen, T., Sharma, S., and Singh, T. P. (2004) *J. Mol. Biol.* **341,** 829–837

6. CONCLUSIONES

Conclusiones

1. A pesar de carecer de la secuencia del genoma de ninguna serpiente y de disponer de un restringido repertorio de secuencias de proteínas de venenos de sólo unas pocas especies de serpientes, la aplicación de técnicas proteómicas -venómica- permite la caracterización exhaustiva de la composición proteica de los venenos de serpientes de la familia Viperinae.

2. Los proteomas de las especies de serpientes Viperinae caracterizados hasta la fecha están constituidos por isoformas de unas pocas (típicamente 10-12) familias de proteínas, cuya abundancia relativa presenta gran variación interespecífica.

3. Debido a su singular estructura modular (metaloproteasa PII con dominio tipo disintegrina PIII) postulamos que BA-A5 representa un intermediario evolutivo en la ruta de diversificación de una metaloproteasa PIII a una disintegrina larga, posiblemente bitistatina.

4. Postulamos que el mecanismo de diversificación de las disintegrinas incluyó la siguiente sucesión temporal de eventos: deleción del dominio rico en cisteínas, pérdida del enlace entre CysXIII-CysXVI, y aparición del motivo de inhibición de integrinas en el ápice de un bucle móvil.

5. El hecho de que todas las secuencias de cDNA amplificadas que codifican para disintegrinas diméricas pertenezcan al tipo "mensajero corto", sugiere que la pérdida del dominio metaloproteasa constituyó un paso importante del mecanismo de emergencia y diversificación de las disintegrinas dimérica.

6. La existencia de mensajeros que codifican para disintegrinas no expresadas en el veneno podría indicar la existencia de "fondos de

reserva genómicos" de eventual relevancia para la adaptación a ecosistemas cambiantes.

7. La disintegrina corta ocelatusina se originó a partir de la transformación de un precursor del grupo de las subunidades de disintegrinas diméricas. Las mutaciones Cys->Tyr y Ser->Cys representan los cambios aminoacídicos mínimos necesarios para convertir una subunidad de disintegrina dimérica en una disintegrina corta.

8. La pérdida sucesiva de intrones forma parte de un mecanismo evolutivo de diversificación de las disintegrinas que conlleva una evolución acelerada y una minimización de las estructuras génica y proteica.

7. BIBLIOGRAFÍA

Alvarez LW, Alvarez W, Asaro F, Michel HV (**1980**) Extraterrestrial cause for the Cretaceous-Tertiary extinction. Science **208**:1095-1108

Arnaout MA, Mahalingam B, Xiong JP (**2005**) Integrin structure, allostery, and bidirectional signaling. Annu Rev Cell Dev Biol **21**:381-410

Bazaa A, Marrakchi N, El Ayeb M, Sanz L, Calvete JJ (**2005**) Snake venomics: comparative analysis of the venom proteomes of the Tunisian snakes *Cerastes cerastes*, *Cerastes vipera* and *Macrovipera lebetina*. Proteomics **5**:4223-35

Becerril B, Marangoni S, Possani LD (**1997**) Toxins and genes isolated from scorpions of the genus Tityus. Toxicon **35**:821-35

Bilgrami S, Tomar S, Yadav S, Kaur P, Kumar J, Jabeen T, Sharma S, Singh TP (**2004**) Crystal structure of schistatin, a disintegrin homodimer from saw-scaled viper (*Echis carinatus*) at 2.5 A resolution. J Mol Biol **341**:829-37

Bilgrami S, Yadav S, Kaur P, Sharma S, Perbandt M, Betzel C, Singh TP (**2005**) Crystal structure of the disintegrin heterodimer from saw-scaled viper (*Echis carinatus*) at 1.9 A resolution. Biochemistry **44**:11058-66

Burgen AS, Roberts GC, Feeney J (**1975**) Binding of flexible ligands to macromolecules. Nature **253**:753-5

Calvete JJ, Schafer W, Soszka T, Lu WQ, Cook JJ, Jameson BA, Niewiarowski S (**1991**) Identification of the disulfide bond pattern in albolabrin, an RGD-containing peptide from the venom

of *Trimeresurus albolabris*: significance for the expression of platelet aggregation inhibitory activity. Biochemistry **30**:5225-9

Calvete JJ, Schrader M, Raida M, McLane MA, Romero A, Niewiarowski S (**1997**) The disulphide bond pattern of bitistatin, a disintegrin isolated from the venom of the viper *Bitis arietans*. FEBS Lett **416**:197-202

Calvete JJ, Moreno-Murciano MP, Sanz L, Jurgens M, Schrader M, Raida M, Benjamin DC, Fox JW (**2000**) The disulfide bond pattern of catrocollastatin C, a disintegrin-like/cysteine-rich protein isolated from *Crotalus atrox* venom. Protein Sci **9**:1365-73

Calvete JJ, Moreno-Murciano MP, Theakston RD, Kisiel DG, Marcinkiewicz C (**2003**) Snake venom disintegrins: novel dimeric disintegrins and structural diversification by disulphide bond engineering. Biochem J **372**:725-34

Calvete JJ (**2004**) Structures of integrin domains and concerted conformational changes in the bidirectional signaling mechanism of αIIbβ3. Exp Biol Med **229**:732-44

Calvete JJ (**2005**) Structure-function correlations of snake venom disintegrins. Curr Pharm Des **11**:829-35

Calvete JJ, Marcinkiewicz C, Monleón D, Esteve V, Celda B, Juarez P, Sanz L (**2005**) Snake venom disintegrins: evolution of structure and function. Toxicon **45**:1063-74

Calvete JJ, Marcinkiewicz C, Sanz L (**2007**) Snake venomics of *Bitis gabonica gabonica*. Protein family composition, subunit organization of venom toxins, and characterization of dimeric disintegrins bitisgabonin-1 and bitisgabonin-2. J Proteome Res (en prensa)

Ching AT, Rocha MM, Paes Leme AF, Pimenta DC, de Fatima DFM, Serrano SM, Ho PL, Junqueira-de-Azevedo IL (**2006**) Some aspects of the venom proteome of the Colubridae snake *Philodryas olfersii* revealed from a Duvernoy's (venom) gland transcriptome. FEBS Lett **580**:4417-22

Clemetson KJ, Lu Q, Clemetson JM (**2005**) Snake C-type lectin-like proteins and platelet receptors. Pathophysiol Haemost Thromb **34**:150-5

Curley GP, Blum H, Humphries MJ (**1999**) Integrin antagonists. Cell Mol Life Sci **56**:427-41

Del Gatto A, Zaccaro L, Grieco P, Novellino E, Zannetti A, Del Vecchio S, Iommelli F, Salvatore M, Pedone C, Saviano M (**2006**) Novel and selective α(v)β3 receptor peptide antagonist: design, synthesis, and biological behavior. J Med Chem **49**:3416-20

Deshimaru M, Ogawa T, Nakashima K, Nobuhisa I, Chijiwa T, Shimohigashi Y, Fukumaki Y, Niwa M, Yamashina I, Hattori S, Ohno M (**1996**) Accelerated evolution of crotalinae snake venom gland serine proteases. FEBS Lett **397**:83-8

Fox JW, Serrano SM (**2005a**) Snake toxins and hemostasis. Toxicon **45**:951-1181

Fox JW, Serrano SM (**2005b**) Structural considerations of the snake venom metalloproteinases, key members of the M12 reprolysin family of metalloproteinases. Toxicon **45**:969-85

Francischetti IM, My-Pham V, Harrison J, Garfield MK, Ribeiro JM (**2004**) *Bitis gabonica* (Gaboon viper) snake venom gland: toward a catalog for the full-length transcripts (cDNA) and proteins. Gene **337**:55-69

Fry BG (**1999**) Structure-function properties of venom components from Australian elapids. Toxicon **37**:11-32

Fry BG, Lumsden NG, Wuster W, Wickramaratna JC, Hodgson WC, Kini RM (**2003**) Isolation of a neurotoxin (alpha-colubritoxin) from a nonvenomous colubrid: evidence for early origin of venom in snakes. J Mol Evol **57**:446-52

Fry BG, Vidal N, Norman JA, Vonk FJ, Scheib H, Ramjan SF, Kuruppu S, Fung K, Hedges SB, Richardson MK, Hodgson WC, Ignjatovic V, Summerhayes R, Kochva E (**2006**) Early evolution of the venom system in lizards and snakes. Nature **439**:584-8

Fukuda K, Mizuno H, Atoda H, Morita T (**2000**) Crystal structure of flavocetin-A, a platelet glycoprotein Ib-binding protein, reveals a novel cyclic tetramer of C-type lectin-like heterodimers. Biochemistry **39**:1915-23

Gottschalk KE, Kessler H (**2002**) The structures of integrins and integrin-ligand complexes: implications for drug design and signal transduction. Angew Chem Int Ed Engl **41**:3767-74

Greene H (**1997**) Snakes: The Evolution of Mystery in Nature. University of California Press, California

Guo M, Teng M, Niu L, Liu Q, Huang Q, Hao Q (**2005**) Crystal structure of the cysteine-rich secretory protein stecrisp reveals that the cysteine-rich domain has a K+ channel inhibitor-like fold. J Biol Chem **280**:12405-12

Gutierrez JM, Rucavado A (**2000**) Snake venom metalloproteinases: their role in the pathogenesis of local tissue damage. Biochimie **82**:841-50

Gutierrez JM, Rucavado A, Escalante T, Diaz C (**2005**) Hemorrhage induced by snake venom metalloproteinases: biochemical and biophysical mechanisms involved in microvessel damage. Toxicon **45**:997-1011

Harrison RA (**2004**) Development of venom toxin-specific antibodies by DNA immunisation: rationale and strategies to improve therapy of viper envenoming. Vaccine **22**:1648-55

Harvey AL (**2001**) Twenty years of dendrotoxins. Toxicon **39**:15-26

Hati R, Mitra P, Sarker S, Bhattacharyya KK (**1999**) Snake venom hemorrhagins. Crit Rev Toxicol **29**:1-19

Hayashi MA, Camargo AC (**2005**) Bradykinin-potentiating peptides from venom gland and brain of *Bothrops jararaca* contain highly site specific inhibitors of the somatic angiotensin-converting enzyme. Toxicon **45**:1163-70

Horii K, Okuda D, Morita T, Mizuno H (**2004**) Crystal structure of EMS16 in complex with the integrin a2-I domain. J Mol Biol **341**:519-27

Huang P, Mackessy SP (**2004**) Biochemical characterization of phospholipase A2 (trimorphin) from the venom of the Sonoran Lyre Snake *Trimorphodon biscutatus lambda* (family Colubridae). Toxicon **44**:27-36

Huang TF, Holt JC, Lukasiewicz H, Niewiarowski S (**1987**) Trigramin. A low molecular weight peptide inhibiting fibrinogen interaction with platelet receptors expressed on glycoprotein IIb-IIIa complex. J Biol Chem **262**:16157-63

Jackson K (**2003**) The evolution of venom-delivery systems in snakes. Zool J Linn Soc **137**:337-354

Jones G, Ronk M, Mori F, Zhang Z (**2001**) Disulfide structure of alfimeprase: a recombinant analog of fibrolase. Protein Sci **10**:1264-7

Joseph R, Pahari S, Hodgson WC, Kini RM (**2004**) Hypotensive agents from snake venoms. Curr Drug Targets Cardiovasc Haematol Disord **4**:437-59

Juárez P, Sanz L, Calvete JJ (**2004**) Snake venomics: characterization of protein families in Sistrurus barbouri venom by cysteine mapping, N-terminal sequencing, and tandem mass spectrometry analysis. Proteomics **4**:327-38

Juárez P, Wagstaff SC, Oliver J, Sanz L, Harrison RA, Calvete JJ (**2006a**) Molecular Cloning of Disintegrin-like Transcript BA-5A from a *Bitis arietans* Venom Gland cDNA Library: A Putative Intermediate in the Evolution of the Long-Chain Disintegrin Bitistatin. J Mol Evol **63**:142-52

Juárez P, Wagstaff SC, Sanz L, Harrison RA, Calvete JJ (**2006b**) Molecular Cloning of *Echis ocellatus* Disintegrins Reveals Non-Venom-Secreted Proteins and a Pathway for the Evolution of Ocellatusin. J Mol Evol **63**:183-93

Junqueira-de-Azevedo Ide L, Ho PL (**2002**) A survey of gene expression and diversity in the venom glands of the pitviper snake *Bothrops insularis* through the generation of expressed sequence tags (ESTs). Gene **299**:279-91

Junqueira-de-Azevedo IL, Ching AT, Carvalho E, Faria F, Nishiyama MY, Jr., Ho PL, Diniz MR (**2006**) *Lachesis muta* (Viperidae) cDNAs reveal diverging pit viper molecules and scaffolds typical of cobra (Elapidae) venoms: implications for snake toxin repertoire evolution. Genetics **173**:877-89

Kamiguti AS, Hay CR, Zuzel M (**1996**) Inhibition of collagen-induced platelet aggregation as the result of cleavage of α2β1-integrin by the snake venom metalloproteinase jararhagin. Biochem J **320**:635-41

Kashima S, Roberto PG, Soares AM, Astolfi-Filho S, Pereira JO, Giuliati S, Faria M, Jr., Xavier MA, Fontes MR, Giglio JR, Franca SC (**2004**) Analysis of *Bothrops jararacussu* venomous gland transcriptome focusing on structural and functional aspects: I--gene expression profile of highly expressed phospholipases A2. Biochimie **86**:211-9

Kini RM, Evans HJ (**1992**) Structural domains in venom proteins: evidence that metalloproteinases and nonenzymatic platelet aggregation inhibitors (disintegrins) from snake venoms are derived by proteolysis from a common precursor. Toxicon **30**:265-93

Kini RM (**2004**) Platelet aggregation and exogenous factors from animal sources. Curr Drug Targets Cardiovasc Haematol Disord **4**:301-25

Kini RM (**2005a**) The intriguing world of prothrombin activators from snake venom. Toxicon **45**:1133-45

Kini RM (**2005b**) Serine proteases affecting blood coagulation and fibrinolysis from snake venoms. Pathophysiol Haemost Thromb **34**:200-4

Kini RM (**2005c**) Structure-function relationships and mechanism of anticoagulant phospholipase A2 enzymes from snake venoms. Toxicon **45**:1147-61

Kini RM (**2006**) Anticoagulant proteins from snake venoms: structure, function and mechanism. Biochem J **397**:377-87

Kisiel DG, Calvete JJ, Katzhendler J, Fertala A, Lazarovici P, Marcinkiewicz C (**2004**) Structural determinants of the selectivity of KTS-disintegrins for the α1β1 integrin. FEBS Lett **577**:478-82

Kochva E (**1987**) The origin of snakes and evolution of the venom apparatus. Toxicon **25**:65-106

Kordis D, Krizaj I, Gunensek F (**2002**) Functional diversification of animal toxins by adaptative evolution. John Wiley & Sons, Chichester, UK

Lu Q, Navdaev A, Clemetson JM, Clemetson KJ (**2005a**) Snake venom C-type lectins interacting with platelet receptors. Structure-function relationships and effects on haemostasis. Toxicon **45**:1089-98

Lu X, Lu D, Scully MF, Kakkar VV (**2005b**) Snake venom metalloproteinase containing a disintegrin-like domain, its structure-activity relationships at interacting with integrins. Curr Med Chem Cardiovasc Hematol Agents **3**:249-60

Marcinkiewicz C, Vijay-Kumar S, McLane MA, Niewiarowski S (**1997**) Significance of RGD loop and C-terminal domain of echistatin for recognition of αIIbβ3 and α(v)β3 integrins and expression of ligand-induced binding site. Blood **90**:1565-75

Marcinkiewicz C, Weinreb PH, Calvete JJ, Kisiel DG, Mousa SA, Tuszynski GP, Lobb RR (**2003**) Obtustatin: a potent selective inhibitor of alpha1beta1 integrin in vitro and angiogenesis in vivo. Cancer Res **63**:2020-3

Marcinkiewicz C (**2005**) Functional characteristic of snake venom disintegrins: potential therapeutic implication. Curr Pharm Des **11**:815-27

Markland FS (**1998**) Snake venoms and the hemostatic system. Toxicon **36**:1749-800

Marsh NA (**1994**) Snake venoms affecting the haemostatic mechanism--a consideration of their mechanisms, practical applications and biological significance. Blood Coagul Fibrinolysis **5**:399-410

McLane MA, Marcinkiewicz C, Vijay-Kumar S, Wierzbicka-Patynowski I, Niewiarowski S (**1998**) Viper venom disintegrins and related molecules. Proc Soc Exp Biol Med **219**:109-19

McLane MA, Sanchez EE, Wong A, Paquette-Straub C, Perez JC (**2004**) Disintegrins. Curr Drug Targets Cardiovasc Haematol Disord **4**:327-55

Menez A (**2002**) Perspectives in Molecular Toxinology, Chichester

Menez A, Stocklin R, Mebs D (**2006**) 'Venomics' or: The venomous systems genome project. Toxicon **47**:255-9

Menezes MC, Furtado MF, Travaglia-Cardoso SR, Camargo AC, Serrano SM (**2006**) Sex-based individual variation of snake venom proteome among eighteen *Bothrops jararaca* siblings. Toxicon **47**:304-12

Mizuno H, Fujimoto Z, Koizumi M, Kano H, Atoda H, Morita T (**1997**) Structure of coagulation factors IX/X-binding protein, a heterodimer of C-type lectin domains. Nat Struct Biol **4**:438-41

Monleón D, Moreno-Murciano MP, Kovacs H, Marcinkiewicz C, Calvete JJ, Celda B (**2003**) Concerted motions of the integrin-binding loop and the C-terminal tail of the non-RGD disintegrin obtustatin. J Biol Chem **278**:45570-6

Monleón D, Esteve V, Kovacs H, Calvete JJ, Celda B (**2005**) Conformation and concerted dynamics of the integrin-binding site and the C-terminal region of echistatin revealed by homonuclear NMR. Biochem J **387**:57-66

Morita T (**2004**) C-type lectin-related proteins from snake venoms. Curr Drug Targets Cardiovasc Haematol Disord **4**:357-73

Morita T (**2005**) Structure-function relationships of C-type lectin-related proteins. Pathophysiol Haemost Thromb **34**:156-9

Moura-da-Silva AM, Theakston RD, Crampton JM (**1996**) Evolution of disintegrin cysteine-rich and mammalian matrix-degrading metalloproteinases: gene duplication and divergence of a common ancestor rather than convergent evolution. J Mol Evol **43**:263-9

Murakami MT, Zela SP, Gava LM, Michelan-Duarte S, Cintra AC, Arni RK (**2003**) Crystal structure of the platelet activator convulxin, a disulfide-linked α4β4 cyclic tetramer from the venom of *Crotalus durissus terrificus*. Biochem Biophys Res Commun **310**:478-82

Niewiarowski S, McLane MA, Kloczewiak M, Stewart GJ (**1994**) Disintegrins and other naturally occurring antagonists of platelet fibrinogen receptors. Semin Hematol **31**:289-300

Ohno M, Menez R, Ogawa T, Danse JM, Shimohigashi Y, Fromen C, Ducancel F, Zinn-Justin S, Le Du MH, Boulain JC, Tamiya T, Menez A (**1998**) Molecular evolution of snake toxins: is the functional diversity of snake toxins associated with a mechanism of accelerated evolution? Prog Nucleic Acid Res Mol Biol **59**:307-64

Ohno M, Ogawa T, Oda-Ueda N, Chijiwa T, Hattori S (**2002**) Accelerated and regional evolution of snake venom gland isozymes. John Wiley & Sons, Chichester, UK

Ohno M, Chijiwa T, Oda-Ueda N, Ogawa T, Hattori S (**2003**) Molecular evolution of myotoxic phospholipases A2 from snake venom. Toxicon **42**:841-54

Olfa KZ, Jose L, Salma D, Amine B, Najet SA, Nicolas A, Maxime L, Raoudha Z, Kamel M, Jacques M, Jean-Marc S, Mohamed el A, Naziha M (**2005**) Lebestatin, a disintegrin from Macrovipera venom, inhibits integrin-mediated cell adhesion, migration and angiogenesis. Lab Invest **85**:1507-16

Olivera BM, Walker C, Cartier GE, Hooper D, Santos AD, Schoenfeld R, Shetty R, Watkins M, Bandyopadhyay P, Hillyard DR (**1999**) Speciation of cone snails and interspecific hyperdivergence of their venom peptides. Potential evolutionary significance of introns. Ann N Y Acad Sci **870**:223-37

Qinghua L, Xiaowei Z, Wei Y, Chenji L, Yijun H, Pengxin Q, Xingwen S, Songnian H, Guangmei Y (**2006**) A catalog for transcripts in the venom gland of the *Agkistrodon acutus*: identification of the toxins potentially involved in coagulopathy. Biochem Biophys Res Commun **341**:522-31

Sanz L, Chen RQ, Perez A, Hilario R, Juárez P, Marcinkiewicz C, Monleón D, Celda B, Xiong YL, Pérez-Payá E, Calvete JJ (**2005**) cDNA cloning and functional expression of jerdostatin, a novel RTS-disintegrin from *Trimeresurus jerdonii* and a specific antagonist of the α1β1 integrin. J Biol Chem **280**:40714-22

Sanz L, Bazaa A, Marrakchi N, Perez A, Chenik M, Bel Lasfer Z, El Ayeb M, Calvete JJ (**2006a**) Molecular cloning of disintegrins from *Cerastes vipera* and *Macrovipera lebetina transmediterranea* venom gland cDNA libraries: insight into the evolution of the snake venom integrin-inhibition system. Biochem J **395**:385-92

Sanz L, Gibbs HL, Mackessy SP, Calvete JJ (**2006b**) Venom proteomes of closely-related Sistrurus rattlesnakes with divergent diets. J Proteome Res **5**:2098-112

Scarborough RM, Rose JW, Hsu MA, Phillips DR, Fried VA, Campbell AM, Nannizzi L, Charo IF (**1991**) Barbourin. A GPIIb-IIIa-specific integrin antagonist from the venom of *Sistrurus m. barbouri*. J Biol Chem **266**:9359-62

Serrano SM, Shannon JD, Wang D, Camargo AC, Fox JW (**2005**) A multifaceted analysis of viperid snake venoms by two-dimensional gel electrophoresis: an approach to understanding venom proteomics. Proteomics **5**:501-10

Shebuski RJ, Ramjit DR, Bencen GH, Polokoff MA (**1989**) Characterization and platelet inhibitory activity of bitistatin, a potent arginine-glycine-aspartic acid-containing peptide from the venom of the viper *Bitis arietans*. J Biol Chem **264**:21550-6

Shimaoka M, Springer TA (**2003**) Therapeutic antagonists and conformational regulation of integrin function. Nat Rev Drug Discov **2**:703-16

Smith JB, Theakston RD, Coelho AL, Barja-Fidalgo C, Calvete JJ, Marcinkiewicz C (**2002**) Characterization of a monomeric disintegrin, ocellatusin, present in the venom of the Nigerian carpet viper, *Echis ocellatus*. FEBS Lett **512**:111-5

Soto JG, Powell RL, Reyes SR, Wolana L, Swanson LJ, Sanchez EE, Perez JC (**2006a**) Genetic variation of a disintegrin gene found in the American copperhead snake (*Agkistrodon contortrix*). Gene **373**:1-7

Soto JG, Powell RL, Reyes SR, Wolana L, Swanson LJ, Sanchez EE, Perez JC (**2006b**) Molecular evolution of PIII-SVMP and RGD disintegrin genes from the genus Crotalus.. Gene (en prensa)

Suto K, Yamazaki Y, Morita T, Mizuno H (**2005**) Crystal structures of novel vascular endothelial growth factors (VEGF) from snake venoms: insight into selective VEGF binding to kinase insert domain-containing receptor but not to fms-like tyrosine kinase-1. J Biol Chem **280**:2126-31

Swenson S, Markland FS, Jr. (**2005**) Snake venom fibrin(ogen)olytic enzymes. Toxicon **45**:1021-39

Takeda S, Igarashi T, Mori H, Araki S (**2006**) Crystal structures of VAP1 reveal ADAMs' MDC domain architecture and its unique C-shaped scaffold. Embo J **25**:2388-96

Theakston RD, Warrell DA, Griffiths E (**2003**) Report of a WHO workshop on the standardization and control of antivenoms. Toxicon **41**:541-57

Wagstaff SC, Harrison RA (**2006**) Venom gland EST analysis of the saw-scaled viper, *Echis ocellatus*, reveals novel α9β1 integrin-binding motifs in venom metalloproteinases and a new group of putative toxins, renin-like aspartic proteases. Gene

Wagstaff SC, Laing GD, Theakston RD, Papaspyridis C, Harrison RA (**2006**) Bioinformatics and Multiepitope DNA Immunization to Design Rational Snake Antivenom. PLoS Med **3**:e184

White JM (**2003**) ADAMs: modulators of cell-cell and cell-matrix interactions. Curr Opin Cell Biol **15**:598-606

White JM (**2005**) Snake venoms and coagulopathy. Toxicon **45**:951-967

Whittaker RH, Margulis L (**1978**) Protist classification and the kingdoms of organisms. Biosystems **10**:3-18

Williams RJ (**1989**) NMR studies of mobility within protein structure. Eur J Biochem **183**:479-97

Xiao T, Takagi J, Coller BS, Wang JH, Springer TA (**2004**) Structural basis for allostery in integrins and binding to fibrinogen-mimetic therapeutics. Nature **432**:59-67

Xiong JP, Stehle T, Zhang R, Joachimiak A, Frech M, Goodman SL, Arnaout MA (**2002**) Crystal structure of the extracellular segment of integrin aVβ3 in complex with an Arg-Gly-Asp ligand. Science **296**:151-5

Yamazaki Y, Hyodo F, Morita T (**2003**) Wide distribution of cysteine-rich secretory proteins in snake venoms: isolation and cloning of novel snake venom cysteine-rich secretory proteins. Arch Biochem Biophys **412**:133-41

Zupunski V, Kordis D, Gubensek F (**2003**) Adaptive evolution in the snake venom Kunitz/BPTI protein family. FEBS Lett **547**:131-6